

Don't rely on Uncle Sam

European regulators should pursue their own investigation into how the 'wrong' genetically modified corn was allowed on the market for years. Unfortunately, their US equivalents show little sign of rising to the challenge.

Given widespread unease among the European public about genetically modified crops, you'd have expected the news that unapproved batches of transgenic seed containing a gene for antibiotic resistance had been shipped from the United States to Europe to have provoked a vigorous official response.

So far, however, reactions from Brussels have been equivocal. Markos Kyprianou, the European commissioner responsible for health and consumer protection, started strongly, declaring that he "deplores" the inadvertent release of unapproved seed in the European Union (EU). But since then, his spokesman has sent mixed messages. First, he simply expressed confidence in the ongoing US investigation into the incident; only later did he state that the European Commission is itself vigorously demanding more information from the company responsible about what happened.

The spokesman's second statement is appropriate; the first seems inadequate. This incident points to fundamental problems with the regulatory framework for agricultural biotechnology in the United States. And the response of the agencies involved gives little confidence that these problems are being seriously addressed.

The facts of the case are these: from 2001 until late last year, a US subsidiary of the Swiss firm Syngenta allowed American farmers to plant 15,000 hectares of corn, or maize, that had been modified with an unapproved version of a gene from the soil bacterium *Bacillus thuringiensis* that codes for an insecticidal protein (see *Nature* **434**, 423; 2005). Small batches of the unapproved seed were also exported to Europe.

After *Nature* revealed this mistake, Syngenta claimed that there was no significant difference between the approved genetically modified corn, called *Bt11*, and the corn that had been inadvertently released, called *Bt10*. Only later did the company admit that *Bt10* differs from *Bt11* in that it contains an additional gene that confers resistance to the antibiotic ampicillin (see *Nature* **434**, 548; 2005) — a difference that most experts agree is of some significance.

System error

Some scientists are shocked that this oversight could have been allowed to persist for so long without detection. One might think that the US federal government, a long-standing champion of agricultural biotechnology, would be hopping mad about the mistake, and keen to get the facts out to satisfy sceptics around the world that it now has the situation firmly under control.

Think again. The US regulatory system divides responsibility between three different agencies, which have collectively failed to respond adequately to this incident. Under a framework introduced during the administration of President Ronald Reagan in the 1980s, the US Department of Agriculture (USDA), the Food and Drug Administration (FDA) and the Environmental Protection Agency (EPA) between them share responsibility for the approval and monitoring of genetically modified crops.

Broadly speaking, the USDA checks to see whether a transgene should be regarded as an agricultural 'pest'; the EPA considers the safety of proteins that can act as pesticides, such as the bacterial toxins expressed by *Bt10* and *Bt11*; and the FDA is responsible for

regulating other food safety aspects of transgenic crops.

The FDA has some justification in taking a back seat in the US investigation, noting that the toxin in *Bt10*, as a pesticide, falls outside its jurisdiction, and judging — quite reasonably — that the antibacterial-resistance gene in *Bt10* does not represent a food safety problem.

The USDA, meanwhile, sees itself primarily as a promoter of US agriculture and related commercial interests — a self-image that has so far been reflected in its handling of this case. Its press office could be taking its line straight from Syngenta: "The system is working," a departmental spokesman said late last month of a process that has taken four years to unearth the cultivation of an unapproved crop. Using the time-honoured technique for burying uncomfortable information, the USDA chose to release news of its decision to fine Syngenta \$375,000 for its error on a Friday afternoon.

Root cause

That leaves the onus on the EPA to investigate the matter. In theory, it has the technical expertise and the legislative authority that it needs to function effectively. But in practice, the EPA's efforts in regulating transgenic crops are mostly devoted to premarketing approvals. According to well-informed critics, the agency has few resources to devote to the *Bt10* investigation. This does little to dispel the message of a story carried on 23 March by the satirical website *The Onion* (www.theonion.com), which suggested that the EPA had announced that it would henceforth be known as The Agency. "We're not really 'environmental' anymore, and we certainly aren't 'protecting' anything," read a quote mischievously attributed to an agency official.

Given the state of the US investigation, the European Commission should establish the relevant facts to its own satisfaction. Syngenta, after all, is a European company — albeit one with headquarters in Switzerland, a non-EU state not known for the transparency of its corporate sector. The company should be forced to reveal how *Bt10* got on to the market in the first place, and why it then took four years to discover the mistake.

So far, we've heard nothing on the former, and Syngenta has attributed the belated discovery of its inadvertent release of *Bt10* to progress in the technology that it uses to monitor seeds. If true, this explanation will come as news to anyone who had assumed that the agricultural-biotechnology industry had known from the start what transgenes it was putting into its seeds. If the discovery was simply a matter of happenstance, we can have little confidence that similar problems won't occur again.

Thankfully, on this occasion we're not dealing with a threat to public health. But the incident will further undermine public confidence. Covering up the circumstances of the current incident may in the short term make life more comfortable for Syngenta and for the regulators who were supposed to prevent the release of unapproved seed. But the long-term prospects for the application of plant genetics would be better served by firm regulation and a degree of candour that has so far been sadly lacking in the performance of the main players in this case. ■



Body of evidence

Anthropologists fear new US law will block access to remains
p810



Calculated risk

Shuttle prepares to launch itself back into active service
p811

Uneasy truce

Palestinian academics question the idea of Israeli collaboration
p813

Safety check

Independent review gives HIV drug clean bill of health
p815

Divergent local laws threaten to stifle Europe's stem-cell project

Quirin Schiermeier, Munich

A high-profile European collaboration in human embryonic stem-cell research is showing signs of trouble before the first pipette has even been unpacked.

Scientists believe that the project could become the continent's most ambitious and promising collaboration in the field so far. But they fear it may be broken up or delayed because of the muddled legal situation surrounding stem-cell research in the European Union (EU).

As *Nature* went to press, representatives of the 25 EU member states were meeting in Brussels to decide whether to approve a €540-million (US\$700-million) package of research projects. The European Commission has already earmarked these for funding under the current Framework Programme on research.

The €12-million stem-cell project is by far the most controversial item on the list. The team behind it includes applicants from Britain, Germany, Sweden, Finland, Italy and France, who plan jointly to investigate the biomedical potential of human embryonic stem cells.

Because such cells can transform into almost any cell type, the hope is that one day they will provide powerful therapies for a wide range of serious conditions, from heart disease to Parkinson's. But isolating them raises ethical concerns, because it requires the destruction of human embryos.

The problem facing the meeting in Brussels is that although a few EU countries,

IMAGE
UNAVAILABLE
FOR COPYRIGHT
REASONS

Ball of contention: Europe remains divided over the ethics of human embryonic stem-cell research.

such as Britain and Sweden, have adopted relatively liberal stem-cell rules, the situation elsewhere is either unclear or markedly more restrictive. Italy, for example, faces a referendum in June about the use of human embryonic stem cells in research, something that Prime Minister Silvio Berlusconi's right-wing government does not support.

And in Germany, most research involving human embryos is illegal, carrying a penalty of up to three years in jail. The use of imported embryonic stem-cell lines created before January 2002 is allowed, provided there is no other way to do the research.

The planned project would involve recent cell lines. So the fear is that EU representatives of countries that don't allow such work will delay the project, or pull out altogether.

"I don't expect it will go smoothly," says

Austin Smith, a stem-cell researcher at the University of Edinburgh, UK, and coordinator of EuroStemCell, the only existing EU-funded stem-cell project, which so far has used mainly adult or mouse cells. "It is still possible that the whole project will be challenged."

Oliver Brüstle, a neuroscientist at the University of Bonn in Germany, hopes to be part of the new collaboration and also appeals to European governments not to block it. "Intellectually, Europe is 100% competitive with the United States," he says. "But it is essential that the strongest groups are able to share their findings and allocate research according to their respective strengths."

Brüstle believes that it would be possible to divide the project up so that German or Italian participants wouldn't come into conflict with their national laws, even when working with colleagues abroad.

Whether or not the proposal pans out, stem-cell research will be overshadowed by a deep political divide between the appointed European Commission and the elected European Parliament. The parliament, traditionally more concerned about the views of individual member countries, recently passed a resolution saying that human embryonic stem-cell research should not get European funding, but be financed from the budgets of countries where such research is legal.

But the commission did not mention this when it announced plans to double the budget for the EU's next Framework Programme on research from 2007 (see page 815). ■

US health officials rally behind bid to relax rules on embryo research

Top officials at the US National Institutes of Health (NIH) have made a pitch for the country's lawmakers to loosen rules that govern federal funding for human embryonic stem-cell research.

The officials lobbied senators who control the agency's budget at a hearing on 6 April. The current policy, set out by President George Bush, prevents scientists from using federal funds to do research on human embryonic stem-cell lines derived after 9 August 2001.

"Access to more cells is seen by scientists as very important to their progress," NIH

director Elias Zerhouni told the hearing.

In written testimony, other NIH leaders were more forceful. "The state of the science is evolving very rapidly and limitations of the president's policy are becoming more apparent," wrote James Battey, director of the National Institute on Deafness and Other Communication Disorders and former head of the NIH Stem Cell Task Force.

Zerhouni has said in the past that scientists should exhaust all the work they can possibly do on permitted stem-cell lines before revisiting the Bush policy. He has also pointed out that the

policy is based not only on science, but on ethics and morality. At the hearing he was careful to state that he did not advocate overturning the policy.

But momentum has been building in Congress to do just that. Scientists have complained loudly that too few lines are available, and that they are contaminated by mouse cells. Support has been growing in the Senate to overturn the Bush rules, and leaders in the House of Representatives last month indicated that they may also be open to revisiting the policy (see *Nature* 434, 548; 2005). **Erika Check, Washington**

Vaccination will work better than culling, say bird flu experts

Declan Butler, Paris

The mass culling of poultry to contain outbreaks of avian flu is no longer acceptable and should be replaced by the vaccination of flocks. That was the conclusion of experts at a joint conference in Paris of the UN Food and Agriculture Organization (FAO) and the World Organisation for Animal Health (OIE).

So far, culling has been the method of choice for controlling outbreaks. Vaccines exist, but the problem has been distinguishing infected birds from vaccinated ones, as both groups carry antibodies to the virus. This means that infected but otherwise healthy, vaccinated birds could provide a haven in which the virus can mutate undetected, increasing the risk of a pandemic strain that causes disease in humans. Trade in vaccinated birds is also subject to bans.

Last year, the OIE and FAO urged that vaccination be considered alongside other control methods. But they have now gone much further, stating that mass culling as the main means of control is no longer acceptable, "for ethical, ecological and economic reasons".

"This is a massive change in policy," says Robert Webster, director of the World Health Organization's Collaborating Center for Studies on the Ecology of Influenza in Animals and Birds.

The shift was driven by a realization that the lethal H5N1 strain is widespread in wild and domestic bird populations, such as ducks. This means that no matter how often affected poultry flocks are culled, the virus is likely to reappear. Methods have also been developed to distinguish infected from uninfected vaccinated birds, either using unvaccinated 'sentinel' birds kept within vaccinated flocks, or vaccines that elicit slightly different antibodies from the natural virus, so that a molecular test can tell the two apart (see *Nature* 427, 573; 2004).

Webster says he is encouraged that the OIE and FAO accept "the reality that H5N1 in Asia is now endemic, and the possibility that more countries will have to turn to vaccination in the longer term".

Experts at the Paris meeting, held on 7–8 April, acknowledged that many countries do not have the resources for a vaccination programme, so they are calling for \$100 million–120 million in aid over three to five years for this purpose. So far, only Germany, Japan and the Netherlands have pledged support for affected Asian countries. ■

IMAGE
UNAVAILABLE
FOR COPYRIGHT
REASONS

Tribes may no longer have to prove a direct cultural link to Kennewick Man to claim his bones.

Law change imperils studies of ancient human remains

Rex Dalton, San Diego

If passed, a bill speeding through Congress could end much research on ancient human remains in the United States. The legislation would alter the Native American Graves Protection and Repatriation Act (NAGPRA), a 1990 federal law that defines how human specimens can be handled, to make it much easier for Native American tribes to claim remains and prevent studies.

The bill was triggered by the legal battle over Kennewick Man, a 9,200-year-old skeleton found in 1996 by the Columbia River in Oregon. Scientists were keen to study it, but tribes in Oregon and Washington wanted to secure the remains so that they could be buried in accordance with traditional beliefs.

Last year, eight scientists working with Oregon-based Friends of America's Past won a long-running legal battle to conduct a comprehensive study of the skeleton. A federal judge ruled that the tribes couldn't show 'cultural affiliation' to Kennewick Man, so they couldn't claim the remains.

The new law is being championed by Senator John McCain, a Republican from Arizona and chairman of the Senate Committee on Indian Affairs, and is designed to sidestep this requirement. His office did not respond to *Nature's* requests for an interview.

The US Senate may vote on the legislation as soon as this week after the measure sailed through the Committee on Indian Affairs on 9 March. If it is passed as expected, the bill will

be considered by the House of Representatives.

Anthropologists are horrified. Cleone Hawkinson, a physical anthropologist who runs Friends of America's Past, said that under the new law, "our prehistory will go behind a black curtain".

Alan Schneider, the scientists' Portland attorney in the Kennewick Man case, explains that the proposed change to the NAGPRA alters the definition of Native Americans from "a tribe, people, or culture that is indigenous to the United States" by adding two words: "or was". That would permit tribes to claim a much wider range of remains even without a direct cultural link to them, he says.

If the bill wins Congressional approval later this year and is signed into law by President George Bush, Schneider says a new legal challenge could be mounted to allow Kennewick Man to be studied. The law would apply to other stored remains, which could be claimed back from museums, as well as newly found specimens. "Even Adam and Eve's remains, if found in this country, would be subject to claims by tribes," he says.

But not all archaeologists agree that the new law will have a major impact. The Society for American Archaeology, whose members primarily focus on artefacts rather than human remains, is not opposing the measure. "This won't disastrously affect research," says Keith Kintigh, an archaeologist at Arizona State University in Tempe, who advises the society on such issues. ■

Shuttle reports for duty...despite the risks

Tony Reichhardt, Washington

Excitement was mounting at NASA last week as a space shuttle rolled onto the launch pad for the first time in more than two years. But although NASA managers are confident that they have done their best to fix the problems that brought down Columbia and its crew in February 2003, the risks facing the 21-year-old shuttle are still significant.

Columbia was doomed after foam insulation from its giant fuel tank hit the vehicle during its ascent to orbit, ripping a hole in the insulating panels on its wing that caused it to burn up as it re-entered Earth's atmosphere. The Columbia Accident Investigation Board, set up after the tragedy, recommended a suite of fixes and additional safety measures before a shuttle could fly again.

Discovery, set for launch from Kennedy Space Center in Florida between 15 May and 3 June, will take seven astronauts to the International Space Station, returning to Earth 12 days later. NASA estimates that, by next year, it will have spent \$1.4 billion on the safety measures, which are focused mainly on preventing the same scenario that caused Columbia's demise. Engineers have redesigned parts of the tank and changed some methods for applying foam. But they still can't guarantee that a small piece of foam might not cause enough damage to prevent a safe re-entry.

So ground-based cameras will track any falling foam pieces, and a laser imaging system and camera attached to the shuttle's robot arm will inspect the vehicle in space. Astronauts on the space station will also



On view: the space shuttle Discovery moves towards its launch pad at Kennedy Space Center.

photograph the shuttle as it turns a slow somersault before docking.

If these inspections detect trouble, NASA has two choices — try to fix the damage in space, or dock Discovery to the space station and send up a second shuttle to bring the astronauts home. Either method carries risks.

Engineers have yet to devise a reliable way of repairing the shuttle in space. On this flight, spacewalking astronauts will test some of the proposed methods. Yet NASA would be reluctant to rely on experimental techniques to fix a real hole in Discovery. Plan B — launching a second, rescue shuttle — has also never been tried.

But shuttle flight director Paul Hill says that two years of studying the details of

insulation damage has made his team better able to make the necessary judgement calls.

And judgement is what flying the shuttle ultimately comes down to. For all the time and money spent, most observers say the risks of shuttle flight are essentially the same as before the Columbia accident.

Ali Mosleh, an expert in reliability engineering at the University of Maryland in College Park, has worked with NASA on a probabilistic risk analysis for the shuttle. "There's no question they have improved aspects of the shuttle system," he says. But "it's unlikely they'll be hit by the same scenario," he points out. Two years on, the risk is "better probably by some percentage," he says. "But not by an order of magnitude." ■

Ideas abound as Japan aims to boost its space image

Ichiko Fuyuno, Tokyo

Japan has launched an ambitious 20-year plan that would rejuvenate its stagnant space programme. Objectives announced by the Japan Aerospace Exploration Agency (JAXA) last week include Earth observation, a Moon probe and a manned space shuttle.

But critics are worried that the vision lacks focus, and that a thinly spread budget could cause problems for the agency.

Public support for JAXA has plummeted in the past couple of years in the wake of some failed missions. New launches were stalled after an H-IIA rocket malfunctioned in November 2003, and a month later the agency lost its Mars probe Nozomi.

JAXA's first long-term plan, announced on 6 April, was well timed, however, coming some six weeks after the successful launch



of an advanced H-IIA rocket.

All the projects in the new plan are "important for the nation and its people," says Ryosuke Futamata of JAXA's strategic planning division. For the first ten years, the main objective is to develop systems for Earth observation and disaster information.

One aim is to give people disaster warnings and evacuation information on their cellphones in the event of an earthquake or tsunami. Another proposal is to develop a system that can measure carbon dioxide emissions in different countries. And space research, including sending probes to the Moon, Mars and Venus, is also a priority.

Most media attention has focused on the part of the plan that includes a manned space shuttle and a base on the Moon. But

this phase faces the most uncertainty — Futamata points out that JAXA won't even ask the government for the money for at least a decade.

Even so, critics are concerned that, with a current annual budget of just ¥180 billion (US\$1.7 billion) — about a tenth of NASA's — JAXA is trying to do too much. Akimasa Sumi, a special member of the government's Space Activities Commission, says that part of the problem is the agency's inability to stop projects once they are under way. This year, five or six satellite launches are planned, with more in the pipeline. "I wonder how these can be funded," Sumi says. "JAXA's plans tend to be pie in the sky."

But Sumi acknowledges that the agency needed a plan that was inspirational, rather than practical. It is thinking big "to win public support amid a backdrop of falling confidence," he says. "It can't say things that sound familiar to people." ■

Health study sets sights on a million people

David Cyranoski and Rachael Williams, Tokyo

Epidemiologists are meeting this week to discuss whether they can pull off the biggest health survey ever attempted. The mammoth project would track the genes, lifestyles and health of more than a million people across Asia, giving researchers unprecedented power to pin down subtle causes of disease.

The aim is to understand how environmental factors such as diet, smoking and exercise affect disease development, and how this risk varies with genetic make-up. The Asian study, dreamed up by John Potter of the Fred Hutchinson Cancer Research Center in Seattle, Washington, is intended to span Malaysia, India, Korea, Japan, China, Taiwan and Singapore. It would dwarf most national efforts, which generally include fewer than

100,000 people. And it would be twice as large, and cover a much wider range of lifestyles, than the European Prospective Investigation into Cancer and Nutrition (EPIC), the largest study of diet and health so far.

Potter plans to focus on people over 50 and follow them for as long as possible. As well as asking questions about lifestyle, the study would collect regular blood samples to monitor how biological markers such as proteins and RNAs change in the early stages of disease.

Some forty epidemiologists and other researchers from Asia and elsewhere are meeting in Seattle this week to discuss the project amid great enthusiasm. "A large sample with varied dietary habits offers major advantages," says Nadia Slimani of the

International Agency for Research on Cancer in Lyon, France, a nutritionist who has worked on EPIC. Single-country studies can miss connections, she says. One wide-ranging European study found a clear link between fibre intake and colon cancer, whereas a similar study based in the United States picked up nothing.

Such a large study would also have the statistical power to show up causes of rare and less-studied diseases such as brain or kidney cancer, says Shoichiro Tsugane of the National Cancer Center in Tokyo, Japan. "It would be very powerful," he says.

And it could turn up uniquely Asian insights into disease. For example, known risk factors for breast cancer, such as few children and obesity, do not hold for the oestrogen receptor-negative class of breast tumours, which are particularly common in Asia. "No one knows what causes these," says Potter.

Pulling off such a huge collaboration won't be easy. Tricky issues include privacy and data distribution, as well as how to compare data from different regions. EPIC got round that last problem by giving standard questionnaires to a small proportion of each study population to calibrate the results.

Working across such diverse cultures will pose special challenges. What foods should go on the questionnaire is just one. Kee Sing Chia, a medical epidemiologist at the National University of Singapore, spent four years working out nutrition tables for just 200 dishes for a health survey of Singapore's Chinese population. Extending that to Singapore's Malay and Indian cuisines and then to others across the region will be a "horrendous task", he says. "Instead of being a researcher, I could open a restaurant." ■

IMAGE
UNAVAILABLE
FOR COPYRIGHT
REASONS

Recruits wanted: the diet and health of a million people across Asia could go under the microscope.

Gene therapists urged to learn more immunology

Erika Check, Washington

How do you resurrect a therapy stalled by a crippling combination of regulatory, financial and scientific hurdles? At a meeting in Arlington, Virginia, last week, gene therapists spelled out the problems in unflinching detail, but concluded that gene therapy can be revived — if its practitioners are willing to make changes.

"This course of events has been experienced by other new therapeutics," said Katherine High, president of the American Society of Gene Therapy, which convened the meeting on 7–8 April. She cited the example of monoclonal antibodies, which went through a cycle of hype, disappointment and eventual medical and commercial success.

Concerns about safety have made patients

reluctant to participate in clinical trials, and regulatory requirements have made trials expensive. The National Institutes of Health (NIH) does not have the resources to fund many clinical trials, and big pharmaceutical companies are not interested in diseases that afflict relatively few patients.

So how can gene therapy continue? Suggestions from the meeting included modifying the molecules used to deliver the genes, developing better animal models, and performing more rigorous safety evaluations before beginning trials in people.

But Daniel Salomon, a transplant surgeon at The Scripps Research Institute in La Jolla, California, who headed up the US Food and Drug Administration's advisory panel on gene therapy until 2003, thinks the field

needs a more profound change of approach.

He argues that trials ran into safety problems because gene therapists did not take the body's immune response seriously enough. "Everything you do that damages healthy tissue, you're going to pay a price for."

Others agree. "Some of us do molecular biology because we don't have to learn immunology," says Savio Woo, a gene therapist at the Mount Sinai School of Medicine in New York.

Salomon is hopeful nevertheless. Gene therapists must understand and plan for the immune system's response to experimental treatments, he advises. This could entail using immunosuppressive drugs. "Gene therapy is going to happen," he says. "A dose of reality is all I'm talking about." ■

IMAGE
UNAVAILABLE
FOR COPYRIGHT
REASONS

Al-Quds University is on the Jerusalem side of the barrier that divides the city from the West Bank.

Palestinian unease sparks fresh calls for Israeli boycott

Jim Giles

For many who follow the Israeli–Palestinian conflict, Al-Quds University in Jerusalem is a beacon of hope. Of all the Palestinian universities, it is the only one that vigorously promotes interaction with Israeli institutions.

But dissent among its academics over such collaborations is becoming increasingly public. A survey of Al-Quds staff, released last month, suggests that most want to avoid joint projects while Israel continues to occupy the West Bank and Gaza.

Although the survey's methodology has come in for criticism, its findings are mirrored in recent Palestinian statements calling for a boycott of Israeli universities — calls that look set to reignite the boycott debate elsewhere.

Al-Quds' policy owes much to its location in a suburb of Jerusalem, where travel to Israeli institutions is not hindered by the Israeli army checkpoints that dot the West Bank and Gaza. Senior staff have backed collaboration for more than a decade, bringing in millions of dollars of funding. At least 70 such projects are ongoing, says Hasan Dweik, the university's acting president.

But according to the recent survey, conducted by the university's employees' union, three-quarters of the university staff oppose collaborative projects. Al-Quds academics, who spoke to *Nature* on condition of anonymity, say they fear that by working with Israelis they are acquiescing to the occupation. "The projects are harmful because they give the impression that the best minds are working together and things are fine," says one faculty member.

Dweik is sympathetic to such fears, but argues that working with Israelis — who are better funded and often more experienced

than their Palestinian colleagues — helps Al-Quds prepare for its role in a future Palestinian state. And, he says, it means that Israeli academics learn about Palestinian views. The survey results, he adds, are contradicted by researchers' actual behaviour: "More than 150 of our 300 academics are involved in joint projects, so how did the union get its result?"

Al-Quds staff who back the survey's findings acknowledged that the questions could have been more rigorously worded. But they point to public opposition among colleagues in the West Bank and Gaza as evidence that the result is representative.

Last July, for example, the Palestinian Campaign for the Academic and Cultural Boycott of Israel was launched. Spearheaded by Lisa Taraki, a sociologist at Birzeit University in the West Bank, the call has been backed by a coalition of academic trade unions from Palestinian universities.

This backing adds momentum to similar campaigns elsewhere, say British academics behind a proposal to boycott Israeli researchers. The UK campaign was launched in 2002 (see *Nature* 425, 444–449; 2003), and later this month Britain's Association of University Teachers will debate a motion to focus the boycott on specific universities, such as those linked with educational institutions in Israeli settlements in the Palestinian territories.

But despite moves from Palestinian academic organizations, those in favour of a boycott will struggle to win over other academics who have already opposed the idea, such as Michael Aizenman, a mathematical physicist at Princeton University in New Jersey. "As a tool for fostering progress in the Middle East, calls for intellectual boycott are highly counterproductive," he says. ■

War of words deepens divide over biodefence funds

Emma Marris, Washington

Relations between disgruntled microbiologists and their paymasters at the US National Institutes of Health (NIH) are degenerating in a very public way.

At issue: whether the influx of cash for studies on potential bioweapons is draining vital resources from basic research. This could leave the United States at far greater risk from natural disease outbreaks than it is ever likely to face from bioterrorism, the researchers fear.

More than 700 NIH-funded microbiologists signed a letter in the 4 March issue of *Science* (307, 1409–1410; 2005). In it they complained that since 2000 the number of grants for non-biodefence-related studies had dropped by 41% for model microorganisms and by 27% for pathogenic microorganisms.

Anthony Fauci, director of the National Institute of Allergy and Infectious Diseases (NIAID), which handles much of the NIH's biodefence work, responded on 1 April with a letter, co-signed by NIH director Elias Zerhouni (*Science* 308, 49; 2005). He painted a rosier picture, saying "from 2000 to 2005, funding for NIAID nonbiodefence research increased by more than 50%".

But this has simply inflamed the organizers of the original complaint. "Seven hundred and fifty people didn't sign this letter because they were confused and there is no issue," says Robert Landick, a bacteriologist at the University of Wisconsin in Madison. And Richard Ebright, a molecular biologist at Rutgers University in Piscataway, New Jersey, calls the reply "frankly dishonest".

So who is right? That depends on how you do the counting. Ebright focused on grants approved by four NIH review panels for work done at several different institutes. Fauci looked for basic microbiology grants across many more panels, including ones for epidemiology, behavioural research and even biodefence, but all within the NIAID.

"The applications are going other places and still getting funded," says John McGowan, the NIAID's director of extramural activities.

But researchers who work at the NIH are unhappy too. A group led by Michael Yarmolinsky of the National Cancer Institute has voiced similar complaints in a closed letter to institute bosses.

The relevant directors say that they will meet both groups to discuss the issue, probably in May. But overcoming the bitter divisions will not be easy. ■

Tough ethics rules at NIH force more staff resignations

Washington The strict new ethics rules at the US National Institutes of Health (NIH) are continuing to cause defections among senior staff.

James Battey, head of the National Institute on Deafness and Other Communication Disorders and former leader of a pivotal stem-cell task force, said on 31 March that he planned to resign. He says new prohibitions on holding stock in biomedical companies clashed with his role as manager of his family's trust fund.

The recently appointed head of the National Institute of Environmental Health Sciences — research physician David Schwartz — has also changed his plans. Schwartz, currently at Duke University in Durham, North Carolina, will not take up his position this month as planned, although he expressed hope that a change in the rules would let him do so in the future.

Even the NIH's director, Elias Zerhouni, is backing off. In a Senate hearing last week, where senators chastised him for going "overboard", Zerhouni admitted that the stock-holding regulations were perhaps too onerous.

Satellites dart towards get-together in space

Washington Two small satellites due to launch this week could have a big impact on space operations.

The US Air Force's XSS-11 satellite, launched on 11 April, will attempt to automatically rendezvous and work with other orbiting satellites during its year-long mission. NASA's DART (Demonstration of Autonomous Rendezvous Technology, pictured) craft is to conduct a similar day-long test following its launch on 15 April.

US space engineers want eventually to dock spacecraft in orbit without human intervention, something the Russians have been doing for years. Automatic rendezvous was proposed last year as part of a robotic mission to service the Hubble Space Telescope, but a National Academy of Sciences review panel deemed it too risky.



Agency bows to pressure over pesticide industry cash

Washington Plans to use money from the chemical industry to fund a study into pesticide exposure in children have been shelved by the US Environmental Protection Agency (EPA).

The Children's Environmental Exposure Research Study was to be a three-year investigation of children's contact with commercial pesticides. But it came under harsh criticism last October from environmental activists, in part because of

a \$2 million contribution from the American Chemistry Council, the United States' largest chemical lobby (see *Nature* 432, 6; 2004).

On 6 April, Senator Barbara Boxer (Democrat, California) vowed to block the confirmation of acting EPA administrator Stephen Johnson as full agency chief unless the study was halted.

"Until this programme is cancelled — no ifs, ands or buts — I am putting a hold on this," she said after the hearing.

In a statement issued on 8 April, Johnson announced that he had agreed to cancel the study.

Study gives backing to HIV drug for mothers

Cape Town Doctors should continue to administer the drug nevirapine — used to reduce mother-to-child HIV transmission in poor countries — despite some high-profile attacks on a study into its use, an expert panel says.

Papers published in 1999 and 2003 by the HIVNET 012 study, conducted in Uganda with funding from the US National Institutes of Health (NIH), found nevirapine to be safe and effective. But NIH employees have since accused the trial researchers of

sloppy record-keeping and of failing to inform participants of the risks involved.

The allegations led some experts to fear doctors would stop using nevirapine. But a US Institute of Medicine panel, asked by the NIH to clarify the issue, stated on 7 April that the HIVNET 012 conclusions are valid. Some adverse events were not reported, the panel found, but under-reporting seemed to occur when patients had multiple adverse events, and at least one adverse event was reported in these cases.

Europe frames plans for funding basic research

Munich Long-awaited plans to establish a European funding agency for basic research seem to be coming to fruition.

The European Commission's proposals for the seventh Framework programme of research, due to run between 2007 and 2013, include €1.5 billion (US\$1.9 billion) a year for a new European Research Council. The agency, which will fund novel and risky research ideas by researchers in the 25 states of the European Union (EU), will base awards purely on scientific merit.

The total framework budget will more than double to €68 billion. But this will be spread over seven years — three more than normal to provide better continuity between

programmes. The main priorities for large research projects are information and communication technologies (€11.2 billion), health (€7.4 billion) and transport (€5.3 billion). The European Parliament and Council will consider the proposals in the next few months.

WHO quells fears of resurfacing flu strain

Tokyo Fears that South Korean pigs had been infected by a strain of flu related to the virus that caused the 1918 pandemic have been refuted by the World Health Organization (WHO).

Concerns surfaced last October, when South Korean scientists published sequences of influenza viruses taken from farm pigs. Some analysts said the data contained genes from WSN/33, the strain derived from the virus that killed tens of millions of people and that now exists only in laboratories.

On 5 April, the WHO said that after analysing further samples, three teams found no evidence of the virus. The WHO says that the sequences were probably submitted in error, perhaps due to a mix-up with lab samples of WSN/33. But the scientists at the Chungnam National University in Daejeon who published the samples say that the possibility of contamination is remote.

IMAGE
UNAVAILABLE
FOR COPYRIGHT
REASONS

Nevirapine can cut HIV transmission to babies.



The dustiest place on Earth

Dust clouds can cool the Earth and halt hurricanes. But the world's biggest dust source was until recently a war zone.

Jim Giles joins one of the few research teams to make the trip.

Midday, near Chicha, northern Chad. It's the dustiest time of day in the dustiest place on Earth. A dirty-white haze drifts in through the open windows of the jeep, coating our clothes with powder. Outside, in the 40 °C heat, there's an argument going on.

Adrian Chappell and Charlie Bristow have come a very long way, to a place where no one else wants to be, to have their dispute: the Bodélé depression, a low-lying region on the eastern fringe of the Sahara. Almost nothing grows here, amid the dunes, gravelly ground and the great plumes of dust that blot out the Sun. Even the local desert people avoid the area.

But it's confusion, not discomfort, that is causing Earth scientists Chappell and Bristow to disagree. After three hot and bumpy days of driving from Chad's capital, N'Djamena, they find themselves in a desert with a surprisingly compact, crusty surface. We're covered in dust, yet it's not clear how the very solid ground we're standing on can produce the clouds that surround us.

This simple and surprising observation will be the source of much discussion over the next two weeks. The team of seven scientists — one of the few groups of researchers ever to visit the Bodélé — will see their tents buried by dust storms, become accustomed to digging jeeps out of the sand, and develop a taste for goat stew and roast gazelle — the only fare in this barren land.

Their efforts will also shed light on one of the major players in Earth's climate. Dust affects storm formation and cloud cover, and provides minerals that fertilize the ocean and jungles thousands of kilometres away. "Features like the Bodélé turn out to be of planetary significance," says team member Richard Washington of the University of Oxford, UK. And yet little is known about exactly how much dust is produced, by what mechanism, and how much of an effect it has on the planet. So perhaps it's no surprise that Chappell, of the University of Salford, and Bristow, from Birkbeck College, London, are now troubled by what seems an easy question — just what causes the Bodélé to pump out dust at such a phenomenal rate?

The pair climb back in the jeep and put their debate on hold until dinner that evening. The seven-man team is camping nearby at Chicha, a place that is little more than a name on the map. Sights: one tree — the only one for tens of kilometres — and several pieces of abandoned weaponry. Apparent population: several dung beetles and one yellow snake.

The Bodélé was not always like this. Our camp would once have been submerged under Lake Megachad, a vast expanse of freshwater that stretched across much of the country, supporting numerous human communities on its shores. Six thousand years ago the lake receded south, and the people went with it. Today's travellers sometimes

C. BRISTOW



Choked: the Bodélé is covered in diatomite, which, when eroded, is the source of huge amounts of dust.

stumble across the remains of those long-gone settlements. On one trip we spy a pile of pottery fragments jumbled up with alligator bones, suggesting that an ancient Bodélian made a meal for the reptile (or vice versa). Later we find a human jawbone and a skull protruding from the gravelly ground, and we duly record their location for future study. “These people probably died around six thousand years ago,” says Andrew Warren, an Earth scientist from University College London. “It’s not inconceivable that we are the next people to come by.”

Wind tunnel

To the northeast of the remains lies the geographical feature that makes the Bodélé so dusty: a gap between two mountain ranges that funnels winds onto the chalky-white desert. The ground here is not made up of sand or rock, but diatomite — the ground-up remains of microscopic freshwater creatures that once thrived in Lake Megachad. Judging from satellite photographs, most of the Bodélé is covered with this material. And those same pictures show that the wind kicks up clouds of diatomite dust that can stretch for up to 700 kilometres. Unlike other dust sources on the planet, the area seems to be active year-round, making it the largest single source of dust (see ‘Measuring up’, overleaf). This can make life

unpleasant in nearby N’Djamena, but dust and other airborne particles have much longer-range effects as well.

Only over the past few decades have scientists begun to realize the range of dust’s impact on the climate. They now know that aerosols such as mineral dust can reflect light from the Sun, cooling the land beneath. Particulate matter also encourages cloud formation, which again tends to reflect light back into space. But other aerosols — particularly carbon from fire and industry — absorb sunlight, heating the atmosphere but not the surface.

Combine these effects, and it seems that the overall impact of particulates is to cool the Earth, reining in the temperature increase caused by greenhouse gases. Yet our understanding of these effects is primitive and the uncertainties great — so great, in fact, that climate scientists cannot rule out the possibility that aerosols have an opposite and unwelcome heating effect.

But while dust has increasingly occupied the minds of researchers, no climate scientist has been to the Bodélé before now. It is easy to see why. A few kilometres away from our camp lies an unexploded bomb, its plump body and tail fin slanting upwards from the sand. We also find a Soviet-made jeep

equipped with a rocket launcher, the armour pierced by bullet holes.

Libya and Chad fought for this area during the 1980s, one of many conflicts to have derailed Chad’s development since it gained independence from France in 1960. Even now the area is turbulent. To the north, the Tibesti mountains remain off-limits due to landmines. East of us, fighting in the Darfur region of Sudan has spilled across the border. And Africa’s deserts are often lawless places anyway. When we camped on the way to Chicha, our guides searched for a hollow away from the road — any lights at night could attract the attention of bandits.

Now here, the priorities for the climate researchers are working out which weather conditions produce dust, and how much sunlight that dust reflects. On arrival, Martin Todd of University College London and Samuel Mbainayel of the Department of Water Resources and Meteorology in N’Djamena, erect a weather station close to camp. That’s complemented by a computer-controlled device, which will track the strength of the sunlight reaching the ground.

Up, up and away

Towards the end of the first day, as the heat recedes a little, Washington and his Oxford colleague Sebastian Engelstaedter begin a somewhat lower-tech experiment. The budget for the trip is extraordinarily tight — the main source of funding is a single grant of just £12,000 (US\$22,500) from the Royal Geographical Society in London. So when Washington and Engelstaedter decided to map wind patterns by tracking the motion of helium-filled balloons, they were forced to improvise to cut costs.

Researchers usually follow the balloons by attaching electric bulbs and batteries and tracking them at night. But with more than 100 flights planned, this was beyond the pair’s resources. Instead, Washington and Engelstaedter attach plastic cups containing candles to their balloons — they tested

this system back in Oxford by driving around town holding the cups and candles out of the window of their car, to check that they wouldn’t blow out. It turns out to work reasonably well; they manage to track the balloons for around 10 minutes, about half the time achievable using electric lights.

As someone who has spent the past 15 years of his holidays criss-crossing the Sahara in a 1960s Land Rover, Washington is undaunted by the prospect of spending two weeks at Chicha. But the others are less enamoured by the idea. On our second day in camp, the winds are strong by midday, raising a gritty cloud that obscures the horizon and gets into mouths and eyes. Laptop keyboards have to be covered with clingfilm.

“If researchers could watch the formation of storms and dust clouds, and better forecast whether they’ll collide, hurricane prediction should improve.”



Remote control: the team of researchers camp at Chicha, which has an apparent population of dung beetles and one snake.

And when the temperature hits 40 °C by mid-afternoon, the team is forced to retreat to what the group has dubbed the 'University of the Bodélé': a tent containing camping chairs and a car stereo for entertainment.

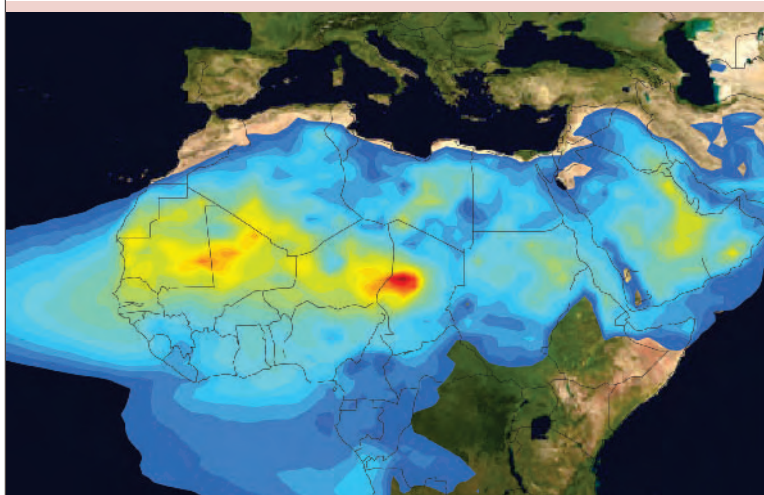
Such weather isn't pleasant, but it's great for data collection. While the researchers huddle in tents, the information being gathered outside will help answer some basic questions with far-reaching impacts. It will determine how the size of the Bodélé dust

cloud varies with wind speed, for example, and how much sunlight the plume reflects. Although satellite pictures have given a good indication of the relative dustiness of this desert, this is no substitute for data gathered from the ground. Armed with this information, modellers will be able to include the conditions and processes special to the Bodélé in their simulations for the first time.

The end result could have dramatic effects on predictions of extreme weather events —

some of them on the other side of the globe. Across the Atlantic in Florida, for example, hurricanes batter the coast almost every summer, running up bills of billions of dollars. The hurricanes form from seedling storms off the west coast of Africa, and over the past few years researchers have discovered that hurricane risk is much reduced if the storm runs into dry and dusty air from the Sahara.

If researchers could watch the formation of seedling storms and dust clouds, and better forecast whether they may collide, hurricane prediction is likely to improve, giving extra days of warning to those on the US east coast. "We're catching hurricanes downstream now," says Jason Dunion, a hurricane researcher at the Atlantic Oceanographic and Meteorological Laboratory in Miami, Florida. "We've got to get to grips with the start of the process."



Dishing the dirt: the Bodélé (centre) is the brightest region in this 13-year average of satellite data on the concentration of aerosols.

Measuring up

The Bodélé's claim to be the dustiest place on Earth rests on data from an instrument originally designed to measure ozone levels: NASA's Total Ozone Mapping Spectrometer (TOMS), launched in 1978.

TOMS measures the amount of radiation that is reflected by Earth's surface and atmosphere at ultraviolet (UV) wavelengths. Different substances in the atmosphere, including dust, reflect varying amounts of UV light at the six different frequencies that TOMS measures. Once this reflectivity signature for aerosols was identified in 1992, TOMS data could be crunched to estimate aerosol concentration.

The aerosols algorithm, developed by a team led by Jay Herman of the Goddard Space Flight Center in Greenbelt, Maryland, has now been applied to all the data generated by the four different satellites that have carried TOMS instruments since 1978. Although other satellite devices have also contributed data on dust

levels, "TOMS is the only one that's truly global", says Martin Todd of University College London.

Use of Herman's algorithm reveals several hotspots. Winds send plumes of mineral dust out westwards from the coasts of Africa and eastwards from China. Land-clearing activities cause fires in equatorial Africa and southeast Asia, which also contribute clouds of smoke. But while these sources wax and wane on an annual cycle, the Bodélé pumps out dust all year round at a rate that matches the peak fluxes seen in other hotspots.

Comparing emissions from different sources is difficult, but, according to one model of global dust distribution, the Sahara produces about two-fifths of global dust emissions (P. Ginoux *et al. Env. Mod. & Software* **19**, 113–128; 2004), whereas the Bodélé is responsible for up to half of all the dust that leaves the west African coast (R. Washington *et al. Ann. Assoc. Am. Geogr.* **93**, 297–313; 2003).

Diatomite dunes

But for the researchers out in the Bodélé, the start of that process is proving perplexing. Before the trip, they had guessed that the Bodélé was covered in talcum-powder fine deposits of diatomite. Washington has seen similar diatomite fields elsewhere in the Sahara and the team assumed that the lightweight particles could easily be whipped up into dust clouds. Wind alone will not do that — smooth air flow will flatten out and then pass over the top of fine deposits — although sand blown in from the surrounding desert could provide the necessary kick.

When we drive out in search of diatomite deposits, armed with Global Positioning System handsets and the satellite imagery as a guide, we find plenty of it, but none of it is in powder form. The white sediments are gravelly in some places, hard and chalky in others. The three Earth scientists find it difficult to fathom how even the most violent sand streams could generate dust. "You'd need an electric sander," says Chappell.

Over that evening's goat stew, the three bat around possible explanations. Bristow suggests that large particles of diatomite might collect in crescent-shaped mounds called barkhan dunes, which march slowly across the desert, blown along a few metres every year. As the wind forces grains up the gentle slope of each dune, from where they



plummet off the sharp edge at the top and down the protruding horns, the particles would be constantly rubbing and grinding against each other. This continual motion might break the chunks of diatomite into particles light enough to blow off the edge of the dune and into the air.

It's a good theory, but a radical one for the field. Most of the world's dust is produced by wind-driven sand bouncing over the top of dusty deposits. And barkhan dunes made from quartz sand are common in deserts, but no one has ever seen one made out of diatomite. As the rest of the camp heads to bed, most sleeping outside to enjoy the cool night, the three are still deep in debate.

But the next day we drive up to a greyish dune and discover that it is not simply coated in a fine layer of diatomite. A sample shows that it seems to be diatomite throughout, made up of chunks of particles of different sizes. "It's extraordinary," says Warren, whose 40-year career has taken him to many Saharan countries. "I don't know of anywhere else with dunes like this." As the wind picks up, we can clearly see that dust is being blown off the crest of these dunes and that the air is clearer where there are fewer dunes.

Chappell and I drive out the following day to set up an experiment that might settle the issue. Around a grey barkhan we set up ten poles, each armed with three bottles to collect sand and dust at different heights. Six

Bottles on poles collect sand and dust blown off the dunes (above) while relics of the 1980s war fought between Chad and Libya remain (below).



poles are planted up the gentle upwind slope of the dune; four are sited off the downwind side. It's sweaty work and Chappell aims to plant 60 in all. Unexpected hitches cause further perspiration: when our vehicle becomes stuck, we shovel sand in the midday sun for an hour to free it.

At first, the weather is on Chappell's side. Thanks to a wind-free period, he plants poles around three dunes and the space around them in just a few days. But when the calm, clear spell continues, he and the others

become worried. Chappell's experiment won't produce much data on quiet days. And the climate scientists want to experience the worst the Bodélé can throw at them, so they can study the weather conditions that generate the huge dust clouds.

Back in London, I check the daily satellite imagery from the comfort of my office. And come the middle of their second week in the Bodélé, a huge plume appears. "It's a beast," Washington tells me by satellite phone. Tents were flattened and covered in sand overnight, sending the inhabitants fleeing to the jeeps at 2 in the morning. The next day, the team had to don ski goggles and masks. Out on the diatomite dunes, it was impossible to distinguish dunes from sky from ground. "It's like flying a plane on instruments," says Chappell.

Flakes and balls

But the data from those last few days, which the team is now analysing back in Britain, will make the trip worthwhile. Washington and Todd are just beginning their analysis, but they are already sure that the wind speeds needed to produce dust in the Bodélé are higher than thought. "Things start getting going at around 12–15 metres per second," says Washington. "A gentle breeze can produce dust elsewhere in the Sahara."

Bristow's theory about dust generation seems to be largely correct. The high winds seen at the end of the trip may release dust as they blow pellets of diatomite across the flat ground, he says, but most of the dust is being generated by flakes and balls of diatomite that grind against each other as the wind pushes them up and over dunes.

Chappell's data provide further confirmation. He sees more dust farther up dunes than at the bottom, and notes that the concentration of dust drops off exponentially with height above the dune tops — a pattern characteristic of a dust source.

Back in his lab in Salford, Chappell has also produced the first estimate of how fast dust flows from the Bodélé. Although the team witnessed just one major storm, the data collected revealed that an average of 40 mg of dust passed every square metre every second during that period — an emission rate that matches the highest recorded anywhere in the world.

So the satellite images that identified the Bodélé as such an important dust source seem to have proven a trusty guide. Those images can now be compared with real data from the ground — and researchers who could not join the trip are eager to get started. "All of us who look at satellite photographs are fascinated with that place," says Joseph Prospero, an expert on dust at the University of Miami in Florida. "We want to know what makes it that way." Thanks to the team from 'the University of the Bodélé' they seem to be getting their wish. ■

Jim Giles is a reporter for *Nature*, based in London.

Science as illusion

When a magician uses science to present his tricks, the effects are seductive.

Alison Abbott takes a masterclass in sorcery.

It seemed rude to leave the lecture hall when the president of the Max Planck Society had generously given the floor to a representative of the “young generation of researchers” in whose hands lie the “future of science”. Still, when Thomas Fraps thanked the society for the opportunity to speak for “55 minutes on some of the many promising themes in science and medicine”, some at the back did quietly slip away.

The remaining audience grew visibly impatient as Fraps warmed to his theme. “My generation is aware of its growing responsibility to bring to the public an informed transfer of new scientific knowledge from the interdisciplinary dialogue within our universities...” They had already sat through a long evening of televised discussion on bioethics, and there was something annoyingly smug, even odd, about this guy. And was his tie really getting longer?

Fraps stepped away from the podium and began self-importantly to clean his glasses with a silk handkerchief from his pocket. Then he pulled the cloth straight through the lenses, flicked the silk to one side and revealed, in his previously empty hand, a glass of orange juice. “Cheers!” he said, taking a sip.

As the audience grasped that Fraps was a magician, astonished laughter erupted. The fact that he used expert scientific patter to accompany his traditional magic tricks added to the delight. A classical rope trick was a demonstration of molecular biology in action, the 2-metre rope representing a strand of DNA. In Fraps’ hands, the two ends mysteriously became four, “to better illustrate the four building blocks of life”, and the rope became a joined circle to illustrate cell division. He used (invisible) ‘genetic scissors’ to snip out a mutated gene. And picking up the small pieces of rope that fell at his feet he recreated the single unbroken length.

The climax of the performance came when Fraps pressed a member of the audience to invest in his research project on transgenic lemons. The ‘volunteer’ handed over a €10 note, after signing it. By sleight of



Casting a spell: Thomas Fraps hovers between the worlds of quantum physics and magic.

hand Fraps turned the note into a worthless receipt. “Well,” he smiled, “if you want good research, you need higher investment.” But he later retrieved it from the dripping centre of a fresh, purportedly transgenic, lemon, cut open on stage. “My research project seems to have been worthwhile after all,” he shouted over resounding cheers.

Fraps earns his living by pretending to be an expert for a range of unsuspecting audiences, not only scientists. Mostly he entertains executives at corporate events for the car or computer industries, where his patter is linked to their professional interests. But his speciality is science-based stage shows, like his current two-hour *Metamagicum* show, centred on quantum physics.

Magic touch

If it hadn’t been for his obsession with magic — he joined the Magic Circle in his native Munich when only 17 — Fraps might have been a conventional physicist. Occasional magic performances helped him pay his way through university, but he didn’t think of turning professional. A short undergraduate stint at CERN, the European particle-physics laboratory near Geneva, taught him that the abstraction of quantum physics was not for him, and he began to think about a career in medical physics. Fatefully, he took a year out before his final exams to do magic — to get rid of the itch once and for all. But the ‘cure’ failed.

Fraps believes that a scientific background can be advantageous for a magician. There are said to be only six conjuring effects,

but hundreds of different ways of achieving them. Magicians mostly make tricks work by directing the audience’s attention away from the object being manipulated. “This is why magic effects can’t be separated from their presentation,” says Fraps.

That presentation must provide engaging storylines with enough logic to make the tricks plausible. “The audience has to believe a trick, so you have to construct a reality where they are going to see cause and effect.” Scientific material makes that easier, he says.

So his earlier struggles with quantum physics have now borne fruit, if not in the context his professors might have expected. “In *Metamagicum*, I transpose a sock sealed in a milk bottle into a parallel universe by using a special machine — the Parallel-O-Meter. In my presentation, this does have internal logic, crazy or not,” he says. Science is weird enough to allow for theoretical parallel universes, he explains, and he simply invents a gadget to do the job.

Fraps sees similarities between Einstein’s ideas of relativity and magic tricks. “The concept of mass or gravity influencing time blew my mind when I was a student,” he says. “Magic blows people’s minds in just the same way, because they see something that contradicts everyday experience.”

Einstein famously said that “the most beautiful thing we can experience is the mysterious — it is the source of all true art and science”. Fraps feels vindicated by this. “My choice to use scientific themes for magic presentation is an artistic one,” he says.

Alison Abbott is *Nature’s* senior European correspondent.

Influenza drug could abort a pandemic

It should be taken, not pre-emptively, but after infection is revealed by a rapid flu test.

Sir—An influenza pandemic will occur at some time in the future: having worked on flu viruses since 1959, I am certain of this. If the deadly H5N1 'bird flu' suddenly acquired human transmissibility, while retaining pathogenicity, the resulting pandemic would cause millions of human deaths. If the pandemic were caused by another bird influenza virus, or if the human H2 virus that disappeared in 1968 were to return, humanity would still be in for a bad time.

No 'pandemic vaccine' could be stockpiled, because of uncertainty about the virus strain. So what can be done? School closure, quarantine, travel restrictions and so on are unlikely to be more effective than a garden hose in a forest fire.

There are, however, two safe antiviral drugs that are effective against all flu viruses, including H5N1. These are the neuraminidase-inhibitors zanamivir/Relenza and oseltamivir/Tamiflu. (Although my crystallization of flu virus neuraminidase led to the development of these drugs and I have a financial interest in Relenza, I have none in Tamiflu.)

Tamiflu is a small carbocyclic molecule that was rationally designed from a knowledge of the X-ray crystal structure of influenza virus neuraminidase. The virus needs neuraminidase to escape from infected cells and spread through the body. The catalytic site of flu neuraminidase, unlike the variable surface antigens, is conserved by all strains of the virus. Tamiflu was designed to fit precisely in the catalytic site of the enzyme, inhibiting its activity.

To be effective, Tamiflu has to be given before infection, or very soon after flu symptoms appear. The time needed to obtain a prescription is a serious drawback.

Although governments around the world are reported to be stockpiling Tamiflu, their strategies for using it are not clear. Britain is reported to have 14.6 million doses of Tamiflu, enough for a quarter of the population. Australia is reported to have enough to protect 200,000 front-line workers prophylactically during a pandemic: two pills a day for 50 days.

This strategy, I believe, is wrong. A more efficient use would be to have Tamiflu available over the counter in local

pharmacies, coupled with a rapid, sensitive and accurate flu diagnostic test. People with flu symptoms could then go immediately to the pharmacy, be rapidly tested to see if the infection is influenza and, if it is positive, be given Tamiflu.

Not only would the infection be curtailed — the person would, on recovery, be immune to reinfection with the same virus.

This procedure could be called 'aborted-infection immunization' and should be used in the early stages of a pandemic, when no vaccine is available, or in the inter-pandemic period by those people who do not take the vaccine or who experience vaccine failure.

The neuraminidase inhibitors exist. They do work. Correct use could be achieved through public education. In a pandemic they would alleviate much of the flu victims' misery, reduce economic losses and probably prevent deaths.

Graeme Laver

Barton Highway, Murrumbateman,
New South Wales 2582, Australia
graeme.laver@bigpond.com

Changes in China call for new health solutions

Sir—The Commentary article "Lessons from the past" by Z. Dong and colleagues (*Nature* **433**, 573–574; 2005), on China's public-health system, has touched upon an urgent issue that may influence the course of China's future. Their attempt to apply such lessons in today's China, however, may require a deeper look at the fundamental changes that have taken place in the country from one era to the other.

The "glorious beginning" under Mao's regime might have been due more to totalitarian control, with communities isolated and movement restricted, than to the health-care system at that time. Although that may have been more fair than today's system, it was at least as inadequate.

The serious medical problems facing today's more open and free China did not have the chance to flourish during that earlier period. Contrast the early 1980s, for example, when sexually transmitted diseases were rare, with the present, when HIV/AIDS has reached epidemic status in parts of the country.

Prevention is the key to public health. But remedies from the past should not be relied on to solve a crisis today, when

one is confronted with totally different problems in a totally different society.

Yonghong Li

Celera Diagnostics, 1401 Harbor Bay Parkway,
Alameda, California 94502, USA

NIH conflicts rules are not right for universities

Sir—We agree with your Editorial "Taking a hard line on conflicts" (*Nature* **433**, 557; 2005) that "scientists and institutions everywhere should be sure that their own houses are fully in order". But the academic research community and the National Institutes of Health (NIH) have different societal roles, expectations and regulatory histories. These differences are essential to understanding why conflict-of-interest policies that apply to researchers employed by the NIH should not be extended to the NIH-funded university community.

Universities' long-standing interactions with industry have produced enormous benefits. Still, the academic community recognizes that the US public's support of academic, and especially biomedical, research depends on maintaining their confidence and trust.

Research universities have had policies on faculty consulting with industry for

decades and have gained deep experience in regulating conflicts of interest across all disciplines. They have given special attention to conflicts in biomedical research since mandatory federal regulations for NIH-funded academics were published in 1995, enhanced by federal guidelines in 2004.

Academia has used these regulations as the base on which to build more robust standards, such as those on human research set out by the Association of American Medical Colleges (AAMC; see www.aamc.org/members/coitf/start.htm). Our 2004 survey of medical schools (see *Nature* **431**, 725; 2004) indicated that most have adopted all or substantial portions of the AAMC recommendations, thereby going well beyond federal requirements.

However stringent the standards and supervision, violations will occur. But to react with overzealous regulations would inhibit useful partnerships and the social benefits that flow from them. As long as universities and medical schools continue to take seriously the enforcement of strict standards, relationships with industry can be principled, protective of research participants and scientific integrity and remain capable of withstanding intense public scrutiny.

David Korn, Susan H. Ehringhaus

Association of American Medical Colleges, 2450
North Street NW, Washington DC 20037, USA

Touching memories

A reminder of the joy and sorrow of reminiscence.

Why Life Speeds Up As You Get Older: How Memory Shapes Our Past

by Douwe Draaisma

(translated by Arnold & Erica Pomerans)

Cambridge University Press: 2004. 288 pp.

£19.99, \$28.99

Yadin Dudai

The science of memory has so far been particularly successful when it has neglected actual memories. The reasons for this have changed over time, yet the paradox remains. Hermann Ebbinghaus, the influential founding father of systematic research into human memory, selected nonsense syllables as study material. This is partly because he wanted to use quantifiable, reproducible stimuli, but also because he was keen to get rid of semantics and to eliminate confounds of associativity in encoding and reconstruction in recall. But real-life memories are all about content — take it away and you are left with mental oblivion.

Animal research didn't fare much better, although it had a more excusable reason: it is inherently more difficult to know what it is like to recollect like a bat than to identify with a recollecting human. The development of simple conditioning paradigms provided an effective way to study performance without having to rely on what the behavioural act means to the brain that commands it. Not surprisingly, it took generations, and thousands of publications, before mainstream animal learning research agreed to consider conditioned animals as knowledge systems, rather than as contentless automata.

On top of all this came the remarkable success of molecular and cellular neurobiology, which often confused plasticity with memory. Memory is specific information about past events, whereas plasticity is believed to be the neuronal property that allows information to be retained over time. Many impressive research articles that claim to deal with memory actually deal with plasticity; their authors seem to have forgotten the distinction.

Some students of human memory prefer to use the term 'memory' only for mental time travel to the personal past, usually involving some re-experienced emotion: I recall, therefore I re-enact. This is bona fide 'episodic' memory. Others take memory to include every type of acquired information that is transparent to conscious awareness, even if this information does not stem from a unique personal experience and there is no mental time travel. This is 'semantic' memory. The textbook adjective for all types

of conscious memory, including both episodic and semantic memory, is 'declarative'; 'non-declarative' refers to all forms of bodily memory, such as habits and modified reflexes, in the performance of which we often behave like zombies.

Humans fear most losing their declarative memories, especially their episodic memory. There is, however, an inverse correlation between the importance we assign to this type of memory in our own brain and our ability to study it in the brains of others. Episodic memory has proved rather resistant to objective investigation. I am not referring here to the current 'neophrenological' wave, in which beautiful false-colour images of brains in action may entice us to believe that we really understand the computational goal of the memory system and the algorithms used to achieve it.

It is refreshing, then, to encounter Douwe Draaisma's book, which reminds us that an interest in memory is primarily synonymous with a wish to understand the joy and sorrow of personal memory. By this we mostly mean memory of personal episodes, which cohere while continuously undergoing both implicit and explicit processing, to yield an autobiographical narrative for the rememberer.

No two brains have identical episodic recollections and autobiographical narratives. These are our unique mental 'brain-prints'. How are these memories formed? Why are some episodes easily recalled, whereas others are forgotten, some only to be recalled later, out of the blue? How can we retrieve complex scenes within a fraction of a second — and why can't we be sure they recapitulate reality faithfully? (They don't.) And why does memory at an advanced age display a bias for recollections from adolescence

(the reminiscence effect or 'bump')?

Draaisma, a historian of psychology at the University of Groningen in the Netherlands, addresses these and other marvels of personal memory in his charming collection of essays. Combining impressive scholarship in the history of psychology and contemporary research with a poetic touch, he discusses the questions outlined above. But he also considers a range of phenomena from exceptional memory and déjà vu to near-death experiences — are we really granted a fast panoramic view of our life seconds before we depart from this world?

The reader should not expect to find here a systematic account of the field of episodic memory, nor extensive coverage of the literature. Rather, this is a fine collection for memory lovers who will appreciate the facts it contains as well as the rich metaphors, which were to be expected in the light of Draaisma's previous book, *Metaphors of Memory* (Cambridge University Press, 2000).

Having said that, even memory researchers might find it a useful reminder of phenomena that have drifted away from the mainstream of the collective scientific memory. The chapter 'Why do we remember forwards and not backwards?' is just one example. I found myself going to the original paper (Bradley, F. H. *Mind* 12, 579–82; 1887) and was soon discussing it at a research seminar. Introspection — a research methodology that seems to have recently regained some of its former glory — tells us that we do indeed recall forwards and not backwards, but this holds only for short episodes. Can time's arrow be the defining attribute of a core episodic item, or 'episodic quantum'? Is there a dedicated brain circuit that subserves this temporal tagging? If so, how is it done

IMAGE
UNAVAILABLE
FOR COPYRIGHT
REASONS

Can it be that it was all so simple then? Or has time rewritten every line?

and are there pathologies that scramble it?

But you don't have to contemplate enigmatic research questions to enjoy this book. Some essays are touching indeed. In 'Why life speeds up as you get older' (which gives the book its title), Draaisma cites a letter from a patient suffering from Alzheimer's disease who no longer remembers that her husband died eight years ago, and keeps writing him heartbreaking letters. Reading this left me not only with a semantic trace, but also with an episodic one. I suspect it will do the same to other readers as well. ■

Yadin Dudai is in the Department of Neurobiology, Weizmann Institute of Science, Rehovot 76100, Israel.

Down to business

University, Inc: The Corporate Corruption of Higher Education
by Jennifer Washburn

Basic Books: 2005. 384 pp. \$26, £19.99

Graham Richards

Money corrupts. It was ever thus, but the thesis of this important book is that the universities' current dependency on industrial links has reached a moral crisis point. Academic freedom, the book contends, has been severely compromised, with damage being done to the public and to the values of higher-education institutions. Here I must declare a conflict of interest, as I have been closely involved in creating spin-off companies and in a multimillion-pound deal to finance a state-of-the-art laboratory. Despite my possible bias in favour of technology transfer, *University Inc.* shakes one's confidence and provokes the realization that some things may have to change.

It is a scholarly book, with almost 700 references and footnotes. However, the author, journalist Jennifer Washburn, cannot resist irritating portraits of many of the major players, for example describing Clark Kerr as "a balding man with bushy white eyebrows and alert blue eyes". It is written exclusively from the United States point of view, where for a long time there has been tension between a utilitarian view of universities, as expressed by Thomas Jefferson, who favoured "a useful American education", and the elitist attitude more prevalent in older European-style institutions, such as Harvard University. Nevertheless, as colleges across the world try to emulate the success of their top US competitors, they are faced with the same dilemmas, encouraged by governments espousing the 'knowledge economy'. Commercial values have been stamped on academic life.

Universities routinely run patenting and licensing operations, invest in risky start-ups, run their own industrial parks and

Exhibition

Passing thoughts



Naglaa Walker, herself a trained physicist, explores the exact world of science from an inexact, intuitive perspective. This photograph captures the image of a passing thought chalked on a blackboard, which is as temporary as the writer wants it to be. It is one of a series in her exhibition *On Physics*, which runs until 7 May at the Hug Gallery for International Photography in Amsterdam.

venture-capital funds, and encourage their faculty to set up companies. But sponsorship brings the risk of ceding control, changing the direction of research and altering the priorities of universities, even to the extent of marginalizing the humanities.

Washburn examines in detail several well-publicized case histories, including an agreement between the University of California, Berkeley, and Novartis. The company gave the university \$25 million over five years for first right to negotiate licences on roughly one-third of the discoveries in the Department of Plant and Microbial Biology. Washburn also takes a rather jaundiced view of the Institute for Bioengineering, Biotechnology and Quantitative Biomedicine (QB3) at Mission Bay, San Francisco — a cooperative effort among three campuses of the University of California and private industry.

The rise of the market-model university dates from the discovery of recombinant DNA technology by Stanley Cohen and Herbert Boyer in 1973 and its subsequent patenting. Washburn provides detailed scrutiny of what she considers the villain of the piece, the Bayh–Dole Act, signed into law by President Carter in December 1980. The bill seems to have been crafted originally to grant

automatic patent rights to just universities and small businesses. It contained safeguards, such as a 'march-in' provision to enable the federal government to terminate licences, and time limits on exclusivity, but few of these survived to the final legislation. Major corporations did not lobby to oppose the act, which the Reagan administration soon extended to include them.

Since the 1980s, the desire to maximize entrepreneurial gains has grown, and they have become significant sources of university funds. DuPont gave \$6 million to Harvard, and the German chemical company Hoechst provided \$50 million for Massachusetts General Hospital — this was targeted money with expected tangible returns in terms of exclusive rights.

The most worrying questions, though, relate to clinical trials. Washburn recounts in detail the problems faced by James Kahn of the University of California, San Francisco, with a trial of the AIDS drug Remune, which was funded by the Immune Response Corporation, and the genetic-engineering trial at the University of Pennsylvania that resulted in the death of Jesse Gelsinger. No one can read these and other case histories in the book without considerable disquiet.

But what is to be done? Pandora's box is already open. We cannot turn back the clock and return to a community of lone scholars, each seeking truth. Modern research in most areas needs massive funding — more than even the richest governments can afford. It is totally unrealistic to place walls between academic institutions and private industry. Universities can, and must, contribute to scientific and technological innovation for wealth creation without compromising their core scholarly principles, academic freedom or essential autonomy. In my opinion, this is becoming even more important as industrial innovations are increasingly the product of small companies, often spun out from universities, while larger corporations concentrate on marketing and selling.

Washburn suggests four fundamental changes. First, she calls for the creation of an independent, third-party licensing body that would assume control over technology-transfer activities in the United States, perhaps allowing some of the more successful campuses to opt out. Second, she has the Bayh–Dole act firmly in her sights, and would have Congress revisit and revise it. In particular, she would change the language so exclusive licensing was the exception and not the norm. Furthermore, she wants the federal government to be able to intercede more easily to provide access to all taxpayer-funded research. The third reform, which would certainly be welcomed in most academic departments, would be to introduce strict conflict-of-interest laws, although the problems in laboratories at the US National Institutes of Health shows how even this might

cause tensions. Finally, she would create a new federal agency to administer and monitor industry-sponsored clinical trials. The fact that there are stirrings in all of these directions suggests that she may well have identified the sensitive areas.

For the world outside the United States—particularly southeast Asia, where research is more related to health and wealth creation—there are lessons to be learned from this book. The financial imperatives need to be kept in check to avoid serious damage, not just to science but to people's lives. ■

Graham Richards is chairman of chemistry at the University of Oxford, and chairman of the intellectual-property company IP2IPO Group.

Politics and history through the mill

Maize and Grace: Africa's Encounter with a New World Crop 1500–2000

by James C. McCann

Harvard University Press: 2005. 320 pp.

\$27.95, £18.95, €25.80

Robert Tripp

Political, social and economic change has often been associated with the introduction and adaptation of new crop species. *Maize and Grace* shows how a New World crop contributed to the emergence of modern-day Africa. Some parts of Africa now have higher maize consumption per capita than Mexico and Guatemala, where the crop originated. Indeed in many African countries, maize production and supply is an issue that attracts an exceptional level of political attention.

It is unlikely that 500 years of history in a continent of diverse cultures and environments could produce a single, consistent narrative of crop adaptation, so James McCann has wisely chosen instead to provide brief glimpses of several important episodes in Africa's engagement with maize. Historical data on the crop's arrival in West Africa and its subsequent use are unfortunately very scanty, and much of the discussion focuses on what linguistic evidence might tell us. More detail is available to discuss the expansion of maize growing (and the priorities of maize breeding) in eastern and southern Africa in the past two centuries. There are also chapters exploring the response of the scientific community to a brief outbreak of rust disease in West Africa in the 1950s, and about a hypothesis linking the recent expansion of maize growing in northwest Ethiopia to a malaria epidemic.

McCann is fairly even-handed in his analysis, avoiding the temptation to cast maize as either hero or villain in the evolution of



Same again? A foreign crop has become Africa's staple.

African agricultural systems. The peasants in this tale are pragmatists, not traditionalists; they are engaged with states and markets, seeking secure livelihoods and ways of coping under different political regimes. The agronomic advantages of maize provided African farmers with a productive early-maturing crop, allowed them to harvest and consume the green ears, and freed them from the bird-scaring required with native cereals. Its economic advantages included wide market demand and suitability for industrial milling techniques.

Maize farming provided opportunities for, and was in turn moulded by, the development of native kingdoms, colonial administrations and independent states in Africa. Crops can, of course, make significant contributions to political regimes without being transferred to new environments: the Incas imposed maize agriculture in their dominion to demonstrate their political control and to feed their armies. And Africa's crops have had an impact in the New World: in the late twentieth century, traditional maize-growing areas in central Mexico were transformed into sorghum fields as the economics of crop production and government policies stimulated the search for alternative crops.

Rather than describing sweeping historical currents, the book offers the reader a series of vignettes that provide opportunities to appreciate the paradoxes of maize development policy and to contemplate some

enduring themes in agricultural history. Development specialists worry about whether Africa's farmers can adapt to increasingly rigid standards for export agriculture. But they may be unaware that the dominance of white maize in southern and eastern Africa owes much to the fact that colonial maize farmers in the early twentieth century planted it to take advantage of export opportunities to Britain, where white starch was in demand. Similarly, current debates about agricultural subsidies in rich countries can be seen in the light of carefully constructed colonial policies that favoured European farmers in Africa.

Investment in agricultural research led to the highly productive single-cross maize hybrid SR52, which helped to support the economy of the breakaway Rhodesian regime. But McCann misses the opportunity to explore how the succeeding Zimbabwe government's devotion to hybrid maize led it to prohibit the sale of open-pollinated varieties.

Although McCann focuses on Africa, he provides some fascinating comparisons and links with southern Europe. During the eighteenth and nineteenth centuries, while maize and cassava were contributing to the growth and political dominance of the kingdom of Asante in West Africa, maize was helping to transform the agriculture of the Venetian republic. And although, as he argues, maize cropping patterns in Ethiopia may be partly responsible for the contemporary spread of malaria, farmers in the Balkans adopted maize to keep them away from the malaria-ridden lowlands. We can mourn the loss of maize diversity in the field and on the table, reflected both in Africans' development of a preference for white maize and in Italians' abandonment of their polenta made from red flint varieties.

The book usefully juxtaposes maize's multiple trajectories in the hands of African farmers with the political and economic forces that increasingly mandate uniform cropping patterns, and with market requirements. These forces are not, of course, unique either to maize or to Africa. McCann fails to answer the question of whether maize can indeed be Africa's "saving grace", and the concluding chapter only touches on the larger questions now facing Africa's agriculture. But the book's nicely drawn examples provide opportunities for reflection as these issues are debated. ■

Robert Tripp is at the Overseas Development Institute, 111 Westminster Bridge Road, London SE1 7JD, UK.

Flight of fancy

How an 'eccentric' line of research proved its worth.

Carol Robinson

From an early age I have been fascinated by the workings of a mass spectrometer. At the age of 16, working in a drug discovery laboratory, I would spend many hours optimizing the mass spectra of small organic chemicals. These volatile molecules were easily vaporized and transmitted into the gas phase, allowing their masses to be determined with high accuracy.

At first I was satisfied with my role as technician, but quickly became frustrated by my lack of scientific background. After many years of part-time study, I graduated in chemistry and later received a PhD in mass spectrometry from Cambridge University. This training stood me in good stead for the developments that took place during my career break. After almost a decade away from science, spent raising my three children, I returned to mass spectrometry and found the same fascination remained. But the mass scale at which I now operated had increased an order of magnitude. During my absence, a momentous innovation had taken place — that of electrospray mass spectrometry. This revolutionary process produced a fine spray from a solution of biomolecules and removed the requirement for volatility. Because of this, it became possible to spray intact proteins into the gas phase of the spectrometer, paving the way for the widespread applications in proteomics that quickly followed this discovery.

I was fortunate to have the opportunity to work with this new technique in the chemistry department at Oxford University. At that time (the early 1990s), the department was using electrospray to characterize small molecules and denatured proteins. I negotiated one day a week on the spectrometer, to indulge in what could be considered quite obscure areas of research. I especially wanted to introduce proteins in conditions where we could maintain their native state in the electrospray droplet, so that I could study the folding process. In the course of this research, I noticed that cofactors and other proteins would sometimes adhere to the folding protein molecule. I became interested in trying to understand this phenomenon, as it had always been assumed that weak interactions between proteins and cofactors would not survive the transition from solution to gas phase.

Despite several early setbacks, I eventually found conditions whereby I could maintain folding and interactions. But my early efforts were restricted by mass range, to complexes containing no more than three or



Critical mass: Carol Robinson pushes the limits of a mass spectrometer to characterize large protein complexes.

four proteins. So I was enthusiastic when invited to look at a prototype mass spectrometer that had an electrospray source interfaced with a time-of-flight detector. I hoped that this device would give me the unlimited mass range I wanted and enable me to look at significantly larger protein complexes. I took some of my most challenging complexes with me, determined to put this spectrometer through its paces and make it work for me.

Unfortunately, my proposal was not greeted with enthusiasm. The instrument had been optimized for high-precision applications, specifically for sequencing peptides. I was seen to want to destroy all the benefits that had been introduced, reducing performance for the sake of obscure research. After several fruitless attempts in which my complexes fell apart into their components, even I was beginning to wonder if this was worth doing. Late one evening, in a last-ditch attempt, we went against traditional mass spectrometry folklore and actively impaired the vacuum system. This action led to success! We were able to observe ions from an assembly of 14 protein subunits of the heat-shock protein GroEL. With a molecular mass of 800,000 daltons, this was the largest protein assembly (at that time) to yield a mass spectrum. Although it was already known that GroEL had 14 subunits, this one spectrum represented a turning point for me and held the promise of many opportunities.

Encouraged by this success and with the beginnings of a research team, we attempted to push the mass range further by projecting into the mass spectrometer intact ribosomes and viruses with masses greater than 2 million daltons, as well as other molecular chaperones. Although the mass spectra of viruses were accepted conceptually (it is well established that their transmission into the atmosphere is all too easy), projecting intact ribosomes into the mass spectrometer was considered more than a little eccentric! I remember, on more than one occasion, being teased by colleagues with suggestions that I was on the 'lunatic fringe'. I was also frequently challenged at conferences. It would be pointed out that the structures of large assemblies in the gas phase were unknown, their formation not understood, and consequently my research was of little significance. Moreover, at that time it was not clear — even to me — what these experiments would show. Examining relatively small molecules provides a unique and identifiable property — that of mass and sometimes molecular composition. For assemblies such as the ribosome, containing upwards of 200,000 atoms, molecular mass is not very informative. We had to find ways of making our unique spectra work for us.

It quickly became apparent that heterogeneity and dynamics were two properties on which we had a new handle. We could measure in real time changes in masses of complexes as they assembled from their components, exchanged subunits between assemblies and changed their folding in response to the addition of factors. These attributes gave us unique insight into protein assemblies that refuse to conform to the requirement of X-ray crystallographers to exist in one rigid form. When working with ribosomes, we discovered that we could examine the dynamic complex known as the 'stalk', which acts like an independent arm in delivering factors to the ribosome. Interestingly — even in high-resolution structures — no individual proteins can be discerned for this part of the ribosome. We were able to uncover both dynamic changes in response to factor binding and a species-dependent stoichiometry.

This short and personal account highlights an interesting aspect of scientific investigation: that the distinction between 'eccentric' and 'progressive' can only be made in hindsight. ■

Carol Robinson is in the Department of Chemistry, University of Cambridge, Lensfield Road, Cambridge CB2 1EW, UK.

Aborting the birth of cancer

Ashok R. Venkitaraman

Can cells sense and stop uncontrolled division driven by cancer-promoting stimuli? Perhaps so, given evidence that aberrant division can trigger the cellular response to DNA damage — blocking growth — at early stages in human cancer.

Why human cancer is not more frequent remains a mystery, given our trillions of susceptible cells, each with many genes subject to mutations that could ignite uncontrolled cell proliferation. One intuitive concept — which has been in the spotlight for decades — is that normal cells can somehow perceive and arrest aberrant cycles of cell division that are triggered by cancer-promoting (oncogenic) stimuli, such as the inappropriate activation of oncogenes. But how cells might do so remains elusive.

On pages 864 and 907 of this issue, Bartkova *et al.*¹ and Gorgoulis *et al.*² supply evidence that oncogene-driven cell-division cycles trigger DNA damage associated with DNA replication (the process that faithfully copies the genome in preparation for division). This DNA damage raises a barrier to sustained proliferation. From these findings, a fresh picture emerges, in which progression towards full-blown cancer requires the wayward cell to inactivate the mechanisms that monitor damage during DNA replication. This would help to explain the close link between genomic instability and cancer evolution, and extend our theoretical framework for understanding how cancers develop.

Early clues to the existence of mechanisms that prevent uncontrolled cell division came from the observation, more than 20 years ago, that viral oncogenes arrest the proliferation of normal cells in culture^{3,4}. Later, the tumour-suppressor proteins p53 and ARF were found to be vital for constraining oncogene-driven proliferation^{5,6}. Their activation was variously attributed to excessive stimulation to proliferate, oxidative stress, or the loss of appropriate signals from the tissue microenvironment⁷ — all triggered by oncogenic stimuli. Activation of these tumour suppressors causes cells either to become dormant (senescence) or to commit suicide (by the process of 'apoptosis'). But evidence that these constraints on proliferation operate during human cancer development has been hard to find.

Enter Bartkova, Gorgoulis and their colleagues^{1,2}, who propose from studies of human cancer samples that another constraint limits aberrant cell division. They provide evidence that the cellular response to DNA damage — specifically, to double-strand breaks in DNA — is activated in early

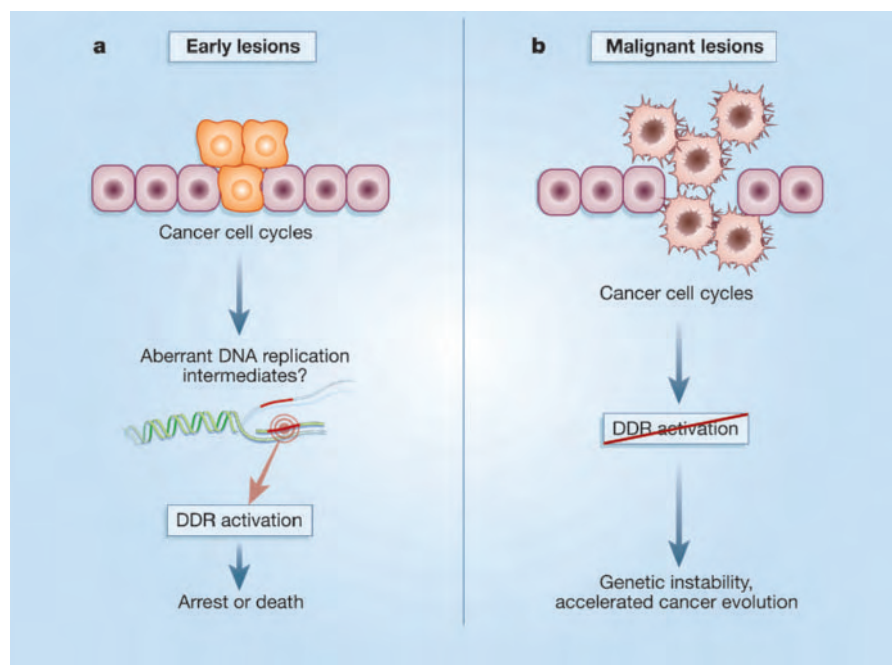


Figure 1 Sensing and stopping wayward cell divisions. a, Bartkova *et al.*¹ and Gorgoulis *et al.*² provide evidence that, in early cancerous lesions, cell-division cycles driven by oncogenic stimuli ('cancer cell cycles') trigger the cellular DNA-damage response (DDR), as a result of aberrations in DNA replication. The nature of these aberrations is uncertain. The DDR then arrests cell proliferation or causes cell death. This might create a selection pressure for suppression of the DDR during carcinogenesis. b, Hence, the progression to malignant lesions might be accompanied by DDR inactivation, which would in turn create genetic instability and accelerate cancer evolution. What distinguishes cancer cell cycles from normal division at the level of DNA replication remains a key, unresolved question.

lesions from lung or bladder tumours. This evidence includes the presence of active forms of ATM or Chk2, participants in the enzymatic cascade that responds to double-strand breaks⁸. Notably, these markers are detected in precancerous lesions — where there is evidence for oncogene-induced aberrant division, but not yet for the changes typical of full-blown cancers — suggesting that the DNA-damage response (DDR) is activated at the earliest stages in carcinogenesis (Fig. 1a). Moreover, the markers are absent from normal proliferating epithelial cells, and from inflammatory lesions, indicating that they discriminate normal from aberrant cell cycles.

To verify and extend these observations, the authors either overexpress oncogenes such as the cell-cycle regulator cyclin E in tissue-culture cells¹, or graft human skin sections onto the backs of immunodeficient mice and use growth factors to induce hyperproliferation of skin cells². In both cases, the

abnormal cell cycles elicit the DDR *in vitro*. This also occurs after inactivation of the tumour-suppressor protein Rb, which ordinarily serves as a gatekeeper for entry into the cell cycle — suggesting that the DDR can be initiated by numerous alterations that underlie the uncontrolled division of cancer cells.

The DDR arrests cell division, and can trigger apoptosis⁹. The authors propose^{1,2} that the need for cells to surmount this barrier during carcinogenesis creates a selection pressure for the inactivation of p53 or other participants in the DDR (Fig. 1b). This, in turn, causes genetic instability — increasing the mutation rate, and accelerating cancer evolution. From this perspective, genetic instability is an unavoidable by-product of the breakdown of barriers to uncontrolled division during early stages of carcinogenesis.

This view raises several questions, the most important of which is how the abnormal

division cycles of precancerous cells — but not equally rapid divisions in normal tissues — can elicit the DDR. The authors argue^{1,2} that the trigger involves DNA replication 'stress': the idea is that the replication machinery performs differently when activated by aberrant, rather than physiological, stimuli. In support of this, the authors offer evidence for abnormalities in replication when proliferation is driven by oncogenic stimuli *in vitro* or in tissues. Provocatively, for example, both groups find that allelic imbalances (signifying chromosomal translocations or deletions) in the genomes of incipient cancer cells occur frequently at 'fragile sites', believed to resist easy copying by the replication machinery.

Replication 'stress' conceived in this way is somewhat nebulous. A hard look at what could distinguish aberrant from physiological stimulation of DNA replication is now needed. Much thinking about the initiation of DNA replication is dominated by a simple 'on-off' concept, in which the replication machinery is loaded onto DNA when key enzymes (cyclin-CDK complexes) are 'off', and then activated when these enzymes are turned 'on'. Yet evidence is emerging¹⁰ for more complex regulatory circuits. Here, the nature and intensity of growth-promoting stimuli are integrated by different levels of activity of cyclin-CDK complexes and of another crucial enzyme, the anaphase-promoting complex, to affect the assembly and operation of the replication machinery. Oncogenic stimuli could perturb these circuits, leading to aberrant replication^{11,12}.

But how aberrant replication might trigger a DDR is far from clear. Cancer cell cycles could generate excessive amounts of normal intermediates, such as single-stranded DNA, spawn abnormal structures, such as double-strand breaks, or even lower the threshold for DDR activation¹³. Identifying the triggering events will be vital to understanding what distinguishes normal from cancer cell cycles.

Alternative scenarios are also possible. In dividing cells, replication is frequently blocked¹⁴ by problems such as oxidative changes to DNA bases. If such problems are unresolved, stalled replication creates abnormal DNA intermediates that trigger the DDR. So, could another difference between normal and cancer cell cycles relate to metabolic changes that augment DNA base lesions? For instance, overexpression of the oncogene *Myc* leads to the production of reactive oxygen species¹⁵. Bartkova *et al.*¹ find that antioxidants have little effect on oncogene-induced DDRs *in vitro*. But, given the high oxygen tensions under which tissue culture is performed relative to *in vivo* conditions, it would be premature to discard this possibility altogether.

Whatever its underlying mechanism,

replication 'stress' as a trigger for DDR activation calls attention to the network of tumour-suppressor proteins that monitors genome integrity during DNA replication¹⁶. Besides ATM, ATR, Chk2 and p53, which enforce cell-cycle checkpoints during the DDR, the network includes Fanconi anaemia proteins and the breast-cancer-susceptibility proteins BRCA1 and BRCA2, which are more directly involved in processing replication-blocking lesions¹⁷. The proposals discussed here suggest that oncogenic stimuli will generate selective pressure for this network to be suppressed during carcinogenesis. Conversely, the proposals could also help to explain why inherited mutations that affect network components, and thereby potentially lower the barrier to uncontrolled division, predispose people to cancer. What we know about the involvement of these tumour suppressors in cancer is not fully consistent with the predictions, however, hinting that further nuances are yet to be discovered.

An interesting twist also reported in this issue is that BRCA2-deficient cells (which cannot deal with stalled replication¹⁸) can be killed by overloading them with replication-blocking DNA damage, using inhibitors of DNA repair^{19,20} (pages 913, 917). Along similar lines, it has been suggested that DDR inhibitors might provide a means to sensitize cancers to therapeutic radiation. The work of Bartkova, Gorgoulis and colleagues^{1,2}

suggests that there could be a long-term price to pay in either situation — in non-malignant cells — if these interventions also overburden, or stifle, the tumour-suppressor network that senses and stops cancer cell cycles.

Ashok R. Venkitaraman is in the Cancer Research UK Department of Oncology and the Medical Research Council Cancer Cell Unit, University of Cambridge, Hills Road, Cambridge CB2 2XZ, UK. e-mail: arv22@cam.ac.uk

1. Bartkova, J. *et al.* *Nature* **434**, 864–870 (2005).
2. Gorgoulis, V. G. *et al.* *Nature* **434**, 907–913 (2005).
3. Tarpley, W. G. & Temin, H. M. *Mol. Cell. Biol.* **4**, 2653–2660 (1984).
4. Franza, B. R. Jr, Maruyama, K., Garrels, J. I. & Ruley, H. E. *Cell* **44**, 409–418 (1986).
5. Kamijo, T. *et al.* *Cell* **91**, 649–659 (1997).
6. Serrano, M., Lin, A. W., McCurrach, M. E., Beach, D. & Lowe, S. W. *Cell* **88**, 593–602 (1997).
7. Lowe, S. W., Cepero, E. & Evan, G. *Nature* **432**, 307–315 (2004).
8. Shiloh, Y. *Nature Rev. Cancer* **3**, 155–168 (2003).
9. Kastan, M. B. & Bartek, J. *Nature* **432**, 316–323 (2004).
10. Diffley, J. F. *Curr. Biol.* **14**, R778–R786 (2004).
11. Ekholm-Reed, S. *et al.* *J. Cell Biol.* **165**, 789–800 (2004).
12. Tanaka, S. & Diffley, J. F. *Genes Dev.* **16**, 2639–2649 (2002).
13. Cox, M. M. *et al.* *Nature* **404**, 37–41 (2000).
14. Shimada, K., Pasero, P. & Gasser, S. M. *Genes Dev.* **16**, 3236–3252 (2002).
15. Vafa, O. *et al.* *Mol. Cell* **9**, 1031–1044 (2002).
16. Osborn, A. J., Elledge, S. J. & Zou, L. *Trends Cell Biol.* **12**, 509–516 (2002).
17. Venkitaraman, A. R. *Nature Rev. Cancer* **4**, 266–276 (2004).
18. Lomonosov, M., Anand, S., Sangrithi, M., Davies, R. & Venkitaraman, A. R. *Genes Dev.* **17**, 3017–3022 (2003).
19. Bryant, H. E. *et al.* *Nature* **434**, 913–917 (2005).
20. Farmer, H. *et al.* *Nature* **434**, 917–921 (2005).

Climate change

Water cycle shifts gear

Thomas F. Stocker and Christoph C. Raible

Various studies indicate that the hydrological cycle is speeding up at high northern latitudes. The resulting increase in freshwater flow into the Arctic Ocean is predicted to have long-range effects.

Discussions of global climate change tend to focus on increasing surface temperatures. By contrast, changes in the water cycle — precipitation, evaporation and river discharge — have received little attention. Yet water has profound effects on our planet's climate. The natural greenhouse effect is caused primarily by water vapour; the radiative balance at the Earth's surface is modified by snow and ice cover; the distribution of vegetation types is sensitive to the local water balance; and regional climate patterns are influenced by ocean currents.

Progress in modelling the many aspects of the water cycle is therefore essential to assess the changes that will result from rising levels of greenhouse gases in the atmosphere. A step forward has been made by Wu and co-workers¹, who, in a study published in *Geophysical Research Letters*, have investigated changes in the freshwater balance of the high northern latitudes.

Wu *et al.*¹ used a climate model that links the influences of the oceans, atmosphere and land surface on climate, and that, for example, has been used to demonstrate the contribution of rising greenhouse gases to warming during the twentieth century². Four simulations modelling the climate over the past 140 years form an ensemble that shows large seasonal cycles and interannual variability. The ensemble gives a figure for mean discharge from Eurasian rivers of about 2.3 Berings (Be; see Fig. 1). This compares with an estimate of 1.9 Be based on observational measurements³. Given the complexity of the hydrological cycle, and the processes that need to be resolved, this discrepancy is remarkably small.

The ensemble suggests an average increase in Arctic river discharge since the mid-1930s of about 1.8 ± 0.6 mBe yr⁻¹, which compares well with the observation-based estimates of 2 ± 0.7 mBe yr⁻¹ (ref. 3).

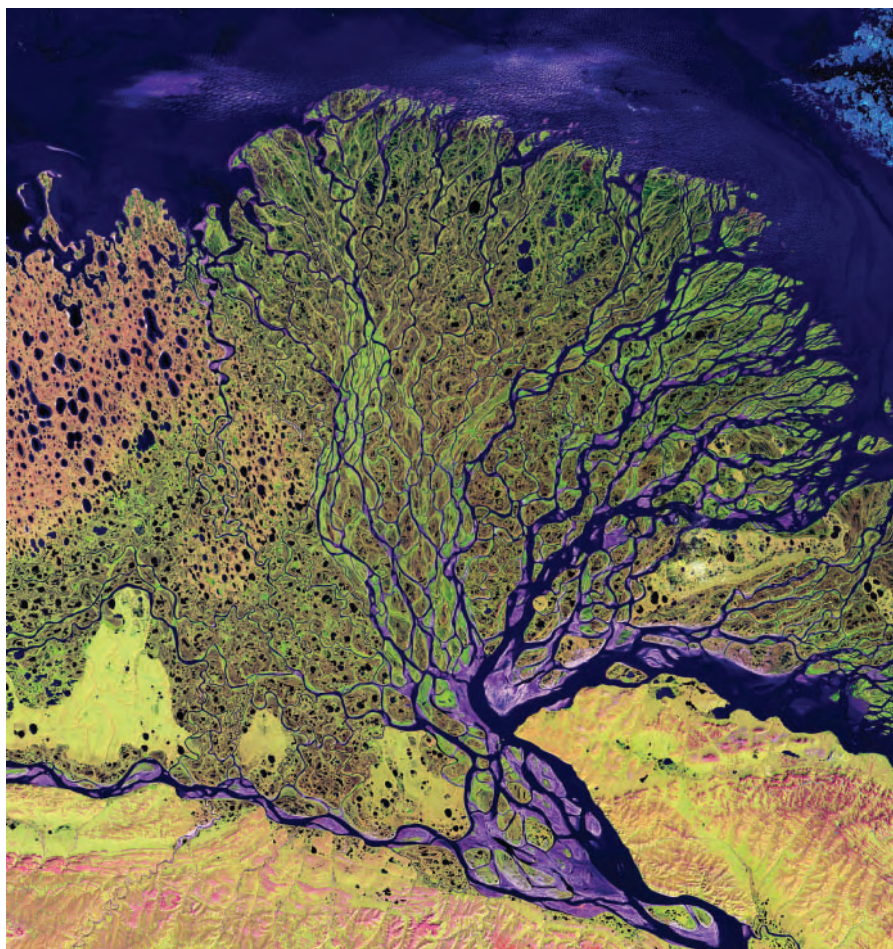


Figure 1 The vast delta of the Lena River, Russia, which is some 250 km wide. Large rivers such as this collect water from Eurasia and northern North America, and discharge it into the Arctic seas. About half of the total volume delivered to the Arctic comes from Eurasian river systems and amounts to about 1.9 Berings; the Bering ($1 \text{ Be} = 10^3 \text{ km}^3 \text{ yr}^{-1}$) is a convenient unit of water flux for large-scale hydrological studies. Like the now popular Sverdrup ($1 \text{ Sv} = 10^6 \text{ m}^3 \text{ s}^{-1}$), used in oceanography, the Bering was proposed by the late Max Dunbar (personal communication). Discharge from Eurasian rivers may increase by 20–70% by the end of this century³.

As in the observations, this trend was magnified by at least a factor of three during the last 30 years of the twentieth century. In the model, a comparison between simulations with and without greenhouse gases reveals that the increased discharge is clearly associated with warming induced by rising concentrations of greenhouse gases. Such trends in water balance are also detectable in the latest data from observations of the atmosphere⁴, particularly in the mid-to-high latitudes of the Northern Hemisphere. The level of natural variability is still large, however, and the changes in the southern high latitudes are less clear because reliable data are not available for the pre-satellite era.

Both observations and models therefore suggest an accelerated hydrological cycle in the atmosphere, which manifests itself as increasing river discharge in the circum-Arctic regions and consequently in a freshening of the northern North Atlantic⁵. Our knowledge of decadal climate variability in the high northern latitudes is limited⁶, and

this variability might affect the fraction of climate change attributable to various causes, particularly with regard to the hydrological cycle. But it seems unlikely that the simulated changes would be quantitatively consistent with natural variability purely by chance.

Next, Wu *et al.*¹ ran their simulations to the end of the twenty-first century, assuming the increased levels of greenhouse gases predicted by two standard emissions scenarios⁷. In these simulations, the acceleration of freshwater delivery to the Arctic, which started in the last three decades of the twentieth century, continues, and by about 2020 the discharge has risen above the upper end of the simulated range of variability. These results reinforce predictions that fundamental environmental changes will occur in the high latitudes of the Northern Hemisphere⁸. The warming will also melt the permafrost and change the seasonal snow cover, both of which will increase the warming and further accelerate the water cycle.

Wu and colleagues' model predicts that

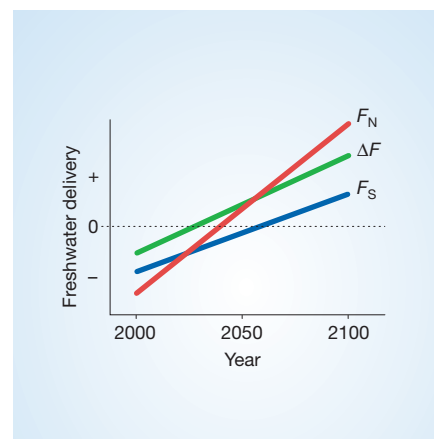


Figure 2 Projected freshwater delivery to the ocean in the high northern (F_N) and high southern (F_S) latitudes. Delivery is projected to increase in both hemispheres, but at a faster pace in the north owing to greater warming. As simulated by the climate model of Wu *et al.*¹, this results in a transfer of fresh water between the hemispheres of $\Delta F = F_N - F_S$. Such a transfer may change the distribution of the major water masses in the world ocean, and thus affect the meridional overturning circulation.

freshwater delivery to the high latitudes will increase in both hemispheres. But the rate of change is larger in the Northern Hemisphere, which will warm more rapidly (Fig. 2). This results in a net transfer of fresh water from the Southern to the Northern Hemisphere. Such a transfer, if it occurs for long enough, could change the properties of the principal water masses in the world's oceans and thus alter the balance of deep-water formation in the far north and far south⁹. Deep-water formation at high latitudes is the process by which warm water, flowing from low latitudes, cools, becomes denser and sinks. The return flow at depth completes the 'meridional overturning circulation'—a major mechanism in oceanic circulation.

As more fresh water is delivered to the Arctic, salinity will be further reduced in areas where deep-water formation takes place, so reducing water density. This might produce a decrease in the meridional overturning circulation in the North Atlantic. Indeed, various models⁷ predict such a reduction, and it is confirmed by the latest climate models, which have a much higher resolution¹⁰.

The range of possible effects is still wide. But it is clear that further and more rapid warming will increase the vulnerability of this circulation system¹¹, possibly leading to a permanent circulation change in the North Atlantic¹². Potentially irreversible changes in the climate system—such as a shutdown of the meridional overturning circulation, or complete melting of the Greenland ice sheet¹³—are now on the radar screens of climate modellers. Such risks will be assessed

in the forthcoming Fourth Assessment Report of the Intergovernmental Panel on Climate Change, due in early 2007.

The evidence for such dramatic changes is still ambiguous, but distributed modelling¹⁴ and coordinated analyses¹⁵ will allow us to improve our estimates of the probability of such events. Results such as those of Wu *et al.*¹ help us on our way. They testify to the ability of today's models to provide a detailed picture of the environmental changes that continued anthropogenic warming of the planet will cause.

Thomas F. Stocker and Christoph C. Raible are at the Division of Climate and Environmental Physics, University of Bern, Sidlerstrasse 5, CH-3012 Bern, Switzerland.

e-mail: stocker@climate.unibe.ch

1. Wu, P., Wood, R. & Stott, P. *Geophys. Res. Lett.* **32**, L02703 (2005).
2. Stott, P. A. *et al.* *Science* **290**, 2133–2137 (2000).
3. Peterson, B. J. *et al.* *Science* **298**, 2171–2173 (2002).
4. Simmons, A. J. & Gibson, J. K. *The ERA40 Project Plan* (European Centre for Medium-Range Weather Forecasts, Reading, UK, 2000).
5. Curry, R., Dickson, B. & Yashayaev, I. *Nature* **426**, 826–829 (2003).
6. Delworth, T. L. & Mann, M. E. *Clim. Dyn.* **16**, 661–676 (2000).
7. Houghton, J. T. *et al.* (eds) *Climate Change 2001: The Scientific Basis* (Cambridge Univ. Press, 2001).
8. Corell, R. *et al.* *Impacts of a Warming Arctic — Arctic Climate Impact Assessment* (Cambridge Univ. Press, 2004).
9. Stocker, T. F., Wright, D. G. & Broecker, W. S. *Paleoceanography* **7**, 529–541 (1992).
10. Bryan, F. O. *et al.* *J. Clim.* (submitted).
11. Stocker, T. F. & Schmittner, A. *Nature* **388**, 862–865 (1997).
12. Knutti, R. & Stocker, T. F. *J. Clim.* **15**, 179–186 (2002).
13. Huybrechts, P., Gregory, J., Janssens, I. & Wild, M. *Earth Planet. Sci. Lett.* **42**, 83–105 (2004).
14. www.climateprediction.net
15. www-pcmdi.llnl.gov

Evolution

Warm-hearted crocs

Adam P. Summers

Our ideas about how crocodiles evolved have just taken a battering. It seems that these cold-blooded creatures, with their limited capacity for prolonged activity, might have had active, warm-blooded ancestors.

A crocodile, quietly submerged, waiting to ambush its next meal, is a testament to the benefits of allowing the environment to determine body temperature. Cool water keeps the metabolic rate low, allowing astonishing periods between breaths that can range into hours. Writing in *Physiological and Biochemical Zoology*, Seymour and colleagues¹ propose that inside these armoured predators beats a heart that harks back to ancestors that were terrestrial, active and — most surprisingly — ‘warm-blooded’.

Warm-blooded, or ‘endothermic’, vertebrates (birds and mammals) have high resting metabolic rates, enabling them to maintain a constant body temperature that is usually above that of their environment. Endotherms use more oxygen and need more fuel than cold-blooded ‘ectothermic’ vertebrates, such as fish, lizards and frogs, with their lower and more variable body temperatures. But, as a consequence, endotherms have gained a capacity for prolonged strenuous activity that can't be matched by ectotherms².

Birds and mammals have independently evolved a four-chambered heart divided into two sides, so that the oxygenated blood coming from the lungs is separated from the deoxygenated blood arriving from the rest of the body (Fig. 1a). The three-chambered heart of most reptiles and amphibians is more flexible, allowing controlled mixing of pulmonary and systemic blood (Fig. 1b). This is especially important in diving, where shunting blood from the lungs accelerates

the transition to a lower metabolic rate, increasing the time between breaths³.

A muscular ridge in the three-chambered heart can completely separate oxygenated from deoxygenated blood, so mere separation of blood flow does not explain why a four-chambered heart is so beneficial that it evolved twice⁴. But in the four-chambered heart, the division of blood flow also allows the pressure to vary between the pulmonary and systemic circulatory systems. The side that sends blood to the lungs is much weaker

than the other side, because high pressures of blood in the lungs could cause fluid to leak across the respiratory membranes. The stronger side of the heart sends blood to the systemic arteries, where the higher pressure works against higher resistance resulting from muscle contraction forces, multiple capillary beds and the extensive filtration across the kidneys. The systemic blood pressures of modern endotherms are three times those of ectotherms.

The crocodilian heart is an interesting mosaic: it is four chambered but it also has a shunting system, although a quite different one from that of other reptiles (Fig. 1c). Seymour *et al.*¹ propose that the ancestors of modern alligators and crocodiles were endothermic and required a four-chambered heart for pressure separation, and that it has since re-evolved a shunting system. The crocodilian ancestors from the Triassic (248 million to 206 million years ago), such as the long-legged *Terrestrisuchus* and the bipedal *Saltoposuchus*, were rather small (less than 2 metres long), completely terrestrial animals that are thought to have had an active lifestyle. They would also have had a diaphragmaticus, a muscle that inflates the lungs by pulling backwards on the liver⁵. This is analogous to the push that our diaphragm muscle exerts on the liver for the same purpose, and it allows higher ventilation rates than contemporary crocodiles require.

In the Jurassic (206 million to 144 million years ago), larger, fully aquatic crocodilian forms developed, and presumably adopted the same sit-and-wait predatory strategy that characterizes modern crocs. Because heat loss in water can be many times higher than that in air, and because larger animals have higher metabolic rates, there would have been considerable selective pressure

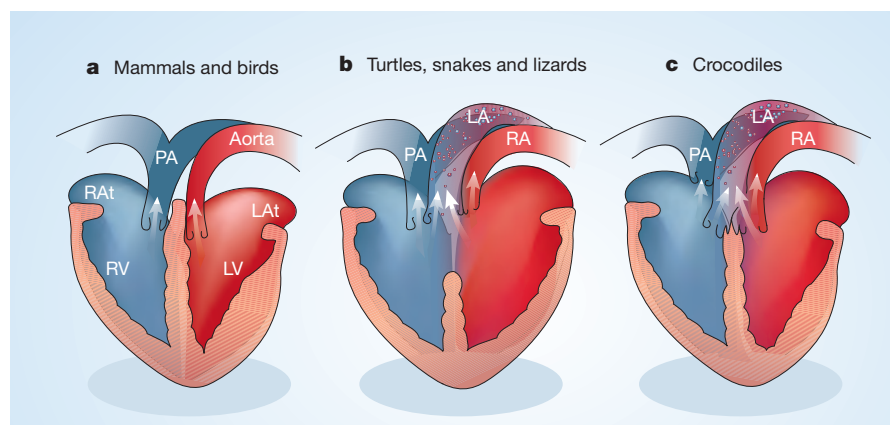


Figure 1 Pump rooms. a, The four-chambered heart of mammals and birds separates oxygenated (red) from deoxygenated (blue) blood, and applies different pressures to the blood that goes to the lungs through the pulmonary artery (PA) and the blood that is pumped through the aorta to the rest of the body. b, The three-chambered heart of most reptiles allows oxygenated and deoxygenated blood to mix and does not permit the large pressure difference between the pulmonary and aortic systems. c, Modern crocodiles have a four-chambered heart, but can ‘shunt’ blood between the two systems through the foramen of Panizza, and are hypothesized¹ to have ancestors that were able to maintain large blood-pressure differences. RA, right atrium; LA, left atrium; RV, right ventricle; LV, left ventricle; RA, right aorta; LA, left aorta.

against endothermy during this return to the water. In addition, animals employing the long dive-times characteristic of an ambush predator would have benefited from a shunting system capable of isolating the lungs or increasing their relative perfusion. DNA extracted from the mitochondria — the cellular structures that use oxygen to convert the potential energy from food into a form the cell can use — provides clues to evolution and energy usage. The mitochondrial DNA from the Jurassic crocodilians shows a high rate of evolution⁶ — roughly equivalent to that of mitochondrial DNA from modern endotherms. This, together with the diaphragmatic muscle and a four-chambered heart, points to an ancestral croc with an elevated metabolic rate.

The idea of meeting an active, agile, hostile croc while hiking is enough to alarm anyone; but the hypothesis of an endothermic crocodilian ancestor has raised comment on several scientific fronts. There are few skeletal characteristics that are convincingly associated with endothermy⁷, but nasal 'turbines' do fit the bill. These scroll-like bones serve as heat and water exchangers in the nose, protecting panting endotherms from desiccation or from getting chilled by large air exchange in the lungs. Crocs have no turbines, and there is no evidence that their fossil ancestors did either. It is also possible that a four-chambered heart is characteristic of an active lifestyle, rather than of endothermy per se⁸.

There is little agreement among experts about the body temperature of dinosaurs, and without significant new fossils direct evidence is unlikely. That leaves the question in the camp of comparative physiologists who study the working of extant animals in hopes of deducing the function of animals long since turned to dust. Evidence from birds and crocodiles, the two living groups that 'bracket' the dinosaurs, is particularly useful. If ancestral crocs had high resting metabolic rates, the clear implication is that dinosaurs did as well.

Adam P. Summers is in the Department of Ecology and Evolutionary Biology, 321 Steinhaus Hall, University of California, Irvine, California 92697-2525, USA.
e-mail: asummers@uci.edu

1. Seymour, R. S., Bennett-Stamper, C. L., Johnston, S. D., Carrier, D. R. & Grigg, G. C. *Physiol. Biochem. Zool.* **77**, 1051–1067 (2004).
2. Bennett, A. F. *J. Exp. Biol.* **160**, 1–23 (1991).
3. Platazack, B. & Hicks, J. W. *Am. J. Physiol.* **281**, R1295–R1301 (2001).
4. Wang, T., Altimiras, J. & Axelsson, M. *J. Exp. Biol.* **205**, 2717–2723 (2003).
5. Carrier, D. R. & Farmer, C. G. *Am. Zool.* **40**, 87–100 (2000).
6. Janke, A. & Arnason, U. *Mol. Biol. Evol.* **14**, 1266–1272 (1997).
7. Bennett, A. F. & Ruben, J. A. in *The Ecology and Biology of Mammal-like Reptiles* (eds Hotton, N., MacLean, P. D., Roth, J. J. & Roth, E. C.) (Smithsonian Institution, Washington DC, 1986).
8. Hillenius, W. J. & Ruben, J. A. *Physiol. Biochem. Zool.* **77**, 1068–1072 (2004).

Geophysics

Double-crossed again

Michael E. Wyssession and Viatcheslav S. Solomatov

An idea that a mineral phase transition may occur not once, but twice, close to the core–mantle boundary has been tested with seismic data. The resulting picture of the deep Earth is sure to provoke debate.

Thermal boundary layers within planets act like thermal bottlenecks, controlling the rate at which heat escapes from the planet and regulating much of its internal behaviour. In the solid Earth, there are two major thermal boundary layers: the lithosphere at the top of Earth's mantle, and the D'' layer at the bottom of the mantle. If we knew the temperature increase across the D'' layer, we would have a sense of the amount of heat flowing out of the core, and thus the power that drives the core geodynamo and the generation of plumes in the mantle.

The temperature gradient across D'' has remained largely unconstrained. However, the model of Hernlund *et al.* (page 882 of this issue)¹ provides a constraint. If the model is correct, the characteristics of the newly discovered mineral phase change from perovskite (Pv), the dominant mineral in the lower mantle) to a post-perovskite (pPv) phase² may have fortuitously given us a means of seismically measuring both the strength and lateral variation of the thermal boundary layer at the base of the mantle.

How does the model work? Phase transitions in Earth's minerals are natural

thermometers, because the pressure at which they happen depends on temperature (this dependence is characterized by the Clapeyron slope). So if we know the pressure, or the depth, at which the phase transition occurs we know the temperature. Hernlund *et al.* push this principle a step further and argue that the Pv–pPv phase transition may happen not just once but twice — perovskite transforms to post-perovskite and then post-perovskite transforms back to perovskite. Earth's temperature profile (geotherm) may therefore 'double-cross' the Pv–pPv phase boundary (Fig. 1a). Such a fortunate situation allows us to estimate not only the temperature, but also the temperature gradient and thus the heat flow out of the core.

So far, this discussion is hypothetical, as we haven't left the laboratory. But this is where we bring in seismology, which serves for geophysics the same role that the telescope serves for astrophysics: it lets you see what's out (or in) there. Each of the mantle's mineral phase changes increases the speed of seismic waves and thus alters the way seismic waves from earthquakes propagate. If the

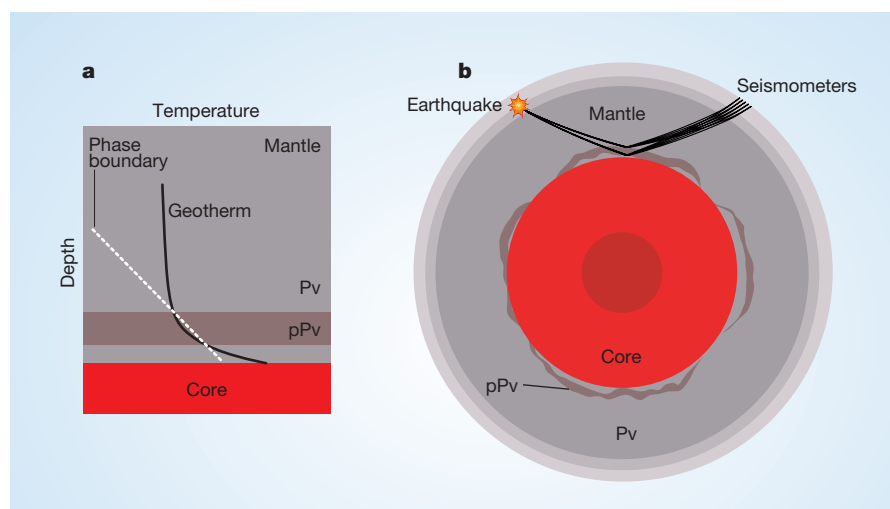


Figure 1 Perovskite (Pv) and post-perovskite (pPv) within Earth's lower mantle. a, Double-crossing of the Pv–pPv phase boundary (dashed line) is inferred from the large Clapeyron slope of the Pv–pPv phase boundary (several times bigger than that of other phase transitions in Earth's mantle) and its location inside the thermal boundary layer where the geotherm changes rapidly. Both crossings create a change in density and seismic velocity that make them 'visible' to seismic waves. b, A hypothetical picture of the global appearance of the post-perovskite layer, based on the model of Hernlund *et al.*¹. Depending on factors such as the local values of the core heat flow and mantle temperature, the Pv–pPv phase boundary may be crossed once (the post-perovskite layer appears all the way to the core–mantle boundary), twice (the post-perovskite layer is sandwiched between layers of perovskite phase) or never (the Pv–pPv transition does not occur at all). A hypothetical set of seismic ray paths is shown sampling both the top and bottom of the Pv–pPv phase boundaries.

phase change is sharp and strong enough, it creates a discontinuity that causes seismic waves to reflect off it (or refract sharply across it). For more than 20 years, seismologists have observed a discontinuity near the base of the mantle that marks the top of the D'' layer. A mineral phase change, including that of the breakdown of perovskite, has long been suggested as a possible cause of the 'D'' discontinuity'. But no particular phase change had been a likely candidate.

However, the discovery last year² of a post-perovskite phase transformation excited seismologists because it occurred at the right depth and had a large enough Clapeyron slope to explain the significant topography observed for the D'' discontinuity. What made the situation even more interesting was the discovery of a newer discontinuity below the D'' discontinuity, one that involves a decrease in seismic wave speeds^{3,4}, but that is distinct from the narrow, ultralow-velocity zone at the core–mantle boundary (which is largely thought to be a zone of partially melted rock). One explanation for the negative discontinuity would be the boundary with a chemically distinct (such as iron-enriched) layer, analogous to the Moho that separates the crust and mantle. But another explanation is that there is a double change in phase at the core–mantle boundary (Fig. 1b).

By finding the depths of the phase boundary crossings, knowing the Clapeyron slope of the phase change and choosing a reasonable value for the thermal conductivity, Hernlund *et al.* are able to estimate the minimum heat flow out of the core. The figure they arrive at is 60–80 mW m⁻²; that is, 9–13 TW for the core as a whole, which is 20–30% of the heat flux from Earth's surface. These estimates are based on a simple linear interpolation. The actual heat flow may be much larger because of the strongly non-linear behaviour of the geotherm in the thermal boundary layer (Fig. 1a). This is more than enough to power the geodynamo and might require an extra heat source within the core, perhaps the radioactive decay of potassium-40. Such a huge heat flow also means that the amount of heat delivered by plumes to the Earth's surface could be much larger than previously thought. Moreover, there may be implications for our understanding of the viscosity and convective patterns within D'': the perovskite beneath the post-perovskite layer might have a lower viscosity and might destabilize the thermal boundary layer, resulting in vigorous internal convection within D''.

This model, of course, will be controversial. Viewing the base of the mantle across nearly 3,000 km of rock is extremely challenging, and there are many competing models for the structure of the lowermost mantle that attempt to explain large anomalies in every seismic parameter imaginable (velocity, attenuation, anisotropy, scattering

and more). The experimental work on the post-perovskite phase is also just beginning.

But it is clear that this is an intriguing model that many geophysicists will be attempting to verify. Old seismic data will be re-examined; other regions of the core–mantle boundary will be investigated; further mineral physics experiments will be carried out to fully examine the mineralogical phase spaces; and the dynamic models of the D'' layer and mantle plumes will have to take into account the new complexities of this enigmatic region. Our understanding of D'' has gone through many revisions over the past 20 years, with this region of Earth

seemingly teasing us with its complications — it is not really surprising that we have been double-crossed again. ■

Michael E. Wysession and Viatcheslav S. Solomatov are in the Department of Earth and Planetary Sciences, Washington University, St Louis, Missouri 63130, USA.

e-mail: michael@mantle.wustl.edu

1. Hernlund, J. W., Thomas, C. & Tackley, P. J. *Nature* **434**, 882–886 (2005).
2. Murakami, M., Hirose, K., Sata, N., Ohishi, Y. & Kawamura, K. *Science* **304**, 855–858 (2004).
3. Thomas, C., Kendall, J. & Lowman, J. *Earth Planet. Sci. Lett.* **225**, 105–113 (2004).
4. Thomas, C., Garnero, E. J. & Lay, T. *J. Geophys. Res.* **109**, doi:10.1029/2004JB003013 (2004).

Neurobiology

Channels for pathfinding

Timothy Gomez

TRP channels are best known for their role in sensory systems: detecting heat and cold, taste, pain and so on. Unexpectedly, they have also been shown to help the growing axons of nerve cells find their way.

The adult nervous system is characterized by a complex network of neurons and their target cells, which are wired up during development when neurons extend processes called axons. The paths taken by axons depend on the movement of their terminal 'growth cones' — structures that control the direction and rate of axon extension in response to molecular cues in the environment. For example, some cues, known as chemotropic factors, can orient the growth cones towards or away from the source of the cue. Growth cones sense external cues through receptors on their surface, and then generate intracellular signals that determine the direction in which the axon grows. One such signal is ionic calcium (Ca²⁺), through which several environmental guidance cues influence the orientation of growth cones. But exactly how these cues promote the appropriate changes in Ca²⁺ levels is poorly defined. On pages 898 and 894 of this issue, Wang and Poo¹ and Li *et al.*² describe an important and unexpected role for a channel protein.

The concentration of intracellular Ca²⁺ in growth cones is tightly controlled by a variety of known channels, pumps and buffers. Fluctuations in intracellular Ca²⁺ concentrations occur when channels in the plasma membrane, or in the membranes of intracellular stores, open to let Ca²⁺ flow into the cytosol. Receptors for external molecular cues can activate channels that raise Ca²⁺ levels in this way. For example, two normally chemoattractive factors — netrin and brain-derived neurotrophic factor (BDNF) — are believed to increase Ca²⁺ levels in neurons by triggering influx of the ion and its release from intracellular stores^{1–3}.

Previous work had suggested that voltage-dependent Ca²⁺ channels (VDCCs) on the plasma membrane are partly, but not completely, responsible for the Ca²⁺ elevations induced by netrin³. However, it was uncertain how netrin depolarizes neurons sufficiently to activate these channels. Wang and Poo¹ and Li *et al.*² now provide conclusive evidence that transient receptor potential (TRP) channels are — either with or without VDCCs — involved in the chemotropic turning of growth cones in response to netrin and BDNF.

TRP channels constitute a diverse family of cation channels that are activated by a variety of means to regulate many biological functions⁴. Although they have been shown to function in nerve growth cones⁵, the new studies^{1,2} are the first to demonstrate that they have a role in chemotropic axon guidance. Together, these studies define a signalling pathway from receptor activation to changes in the concentration of Ca²⁺ in the cytosol that lead to growth-cone turning (Fig. 1, overleaf).

In this pathway, netrin or BDNF first stimulates its receptor (DCC or TrkB, respectively), resulting in the activation of the enzyme phospholipase C. This then produces a messenger molecule, inositol-1,4,5-trisphosphate (IP₃), which in turn causes the release of Ca²⁺ from intracellular stores. Ca²⁺ release from the stores then activates TRP channels on the cell surface, through a poorly understood mechanism. In amphibian spinal neurons¹, the opening of TRP channels (which allow both Ca²⁺ and Na⁺ to enter) depolarizes neurons enough to activate VDCCs, which bring about additional Ca²⁺ influx. However, this amplification of

Ca^{2+} signals by VDCCs may not be universal, as it appears that VDCCs are not involved in Ca^{2+} influx in response to BDNF in mammalian neurons².

So how do chemotropic gradients of netrin or BDNF thereby orient axon growth? As has been shown previously, Wang and Poo¹ and Li *et al.*² find that local application of these guidance cues results in a localized increase in intracellular Ca^{2+} on one side of growth cones, because of local Ca^{2+} release from stores and influx through TRP

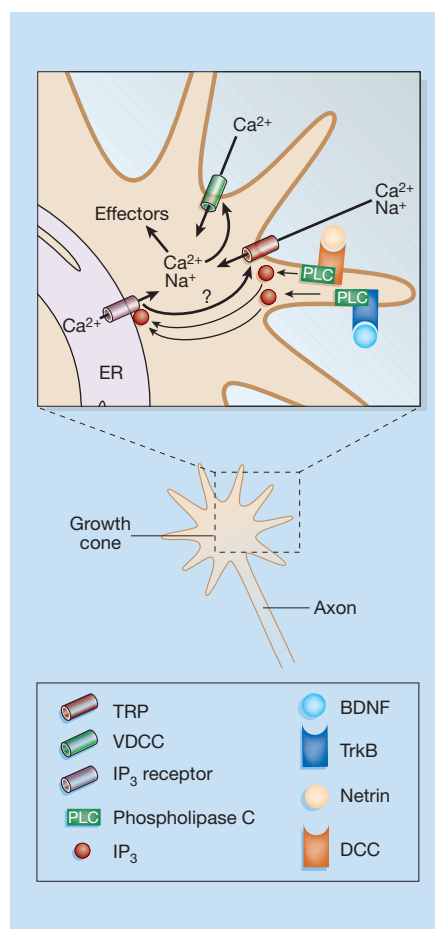


Figure 1 Calcium signals and axon guidance. TrkB and DCC — receptors for two axon-guidance cues, brain-derived neurotrophic factor (BDNF) and netrin, respectively — are believed to activate phospholipase C, which generates inositol-1,4,5-trisphosphate (IP_3) through lipid hydrolysis. IP_3 stimulates Ca^{2+} release from intracellular stores (such as the endoplasmic reticulum, ER) through IP_3 receptors. Wang and Poo¹ and Li *et al.*² show that Ca^{2+} release activates TRP channels, which allow influx of both Ca^{2+} and Na^+ . Influx of Na^+ depolarizes the neuron sufficiently to activate voltage-dependent Ca^{2+} channels (VDCCs), further enhancing Ca^{2+} influx. If these Ca^{2+} signalling pathways are activated disproportionately on one side of a growth cone, as occurs when there is a concentration gradient of guidance cues, local activation of Ca^{2+} -dependent effector molecules will occur, turning the axon in the appropriate direction.

channels. Blocking either release or influx prevents turning, suggesting that both are necessary. This was elegantly demonstrated using a technique known as RNA interference to eliminate a specific subset of TRP channels.

However, this conclusion is slightly ambiguous, because it is often difficult to alter Ca^{2+} release from stores without affecting Ca^{2+} influx, and vice versa. This is certainly true when blocking Ca^{2+} release, which is required to stimulate influx through TRP channels. And it might also be true of blocking TRP-channel activity, which may be required to fill intracellular Ca^{2+} stores. The distinction between Ca^{2+} release and influx is important, as it is uncertain whether both pathways activate similar downstream targets that affect growth-cone motility, leading to turning. Future studies should investigate the specific effects of Ca^{2+} release versus influx on Ca^{2+} -activated molecular targets and the behaviour of growth cones.

With the findings of Wang and Poo¹ and Li *et al.*², the signalling cascade responsible for chemotropic axon guidance has gained an important new player: the TRP channel. Many questions remain unanswered, however. For example, it is not known how the receptors for netrin and BDNF stimulate phospholipase C to produce IP_3 . Further downstream, the targets of the Ca^{2+} signals that are generated by initial Ca^{2+} influx and release are also uncertain, although protein kinase C, CaMKII and calpain^{6–8} are candidates. Processes that are directly responsible for turning growth cones probably ultimately involve local regulation of the cellular skeleton, and perhaps endocytosis or exocytosis (mechanisms that ingest or extrude membrane) as well. These processes all involve proteins of the Rho family. So it is likely that Ca^{2+} signals ultimately modulate the activity of Rho proteins locally to stimulate axon outgrowth up gradients of netrin and BDNF⁶.

Interestingly, under some conditions, netrin and BDNF can also act as chemorepellants, directing growth cones down a concentration gradient. Under these conditions, it has been suggested that a shallower intracellular Ca^{2+} gradient will activate different downstream targets that promote repulsive turning^{7,9}. One possible mechanism that could produce a smaller local Ca^{2+} increase is the modulation of Ca^{2+} currents through VDCCs by cyclic nucleotides (via PKA)¹⁰. However, because VDCCs do not appear to be involved in the chemotropism of mammalian axons, other channels may be modulated instead of these. One possibility that arises from the new findings^{1,2} is that TRP channels might be modulated by cyclic nucleotides to control the amplitude of Ca^{2+} gradients within growth cones. And no doubt other channels and molecular targets await



100 YEARS AGO

“The Dynamical Theory of Gases.” In Mr. Jean’s valuable work on this subject he attacks the celebrated difficulty of reconciling the “law of equipartition of energy” with what is known respecting the specific heats of gases. Considering a gas the molecules of which radiate into empty space, he shows that in an approximately steady state the energy of vibrational modes may bear a negligible ratio to that of translational and rotational modes. I have myself speculated in this direction; but it seems that the difficulty revives when we consider a gas, not radiating into empty space, but bounded by a perfectly reflecting enclosure. There is then nothing of the nature of dissipation; and, indeed, the only effect of the appeal to the aether is to bring in an infinitude of new modes of vibration, each of which, according to the law, should have its full share of the total energy. I cannot give the reference, but I believe that this view of the matter was somewhere expressed, or hinted, by Maxwell. We know that the energy of aetherial vibrations, corresponding to a given volume and temperature, is not infinite or even proportional to the temperature. For some reason the higher modes fail to assert themselves.

RAYLEIGH

From *Nature* 13 April 1905.

50 YEARS AGO

Sir Alexander Fleming, the discoverer of penicillin, died suddenly on March 11 at his home in London... The First World War directed his attention to wound infection and its control; he was one of the first to show that most wounds become infected after admission to hospital... In 1922 came the discovery of lysozyme. Following an observation that his own nasal secretion after a common cold had an inhibitory action on some of the bacteria present in the nose, he showed that this lytic substance was present in many tissues and secretions of the body... His work with lysozyme prepared the way for the epoch-making discovery of penicillin. Fleming had been studying variation in the staphylococcus, which meant that he was frequently examining colonies of that organism as they appeared on ordinary nutrient agar. As a result of this exposure, a mould appeared on the culture medium some days later and Fleming noticed the unusual phenomenon of the disappearance of staphylococcal colonies around the mould.

From *Nature* 16 April 1955.

discovery, given the diversity of Ca^{2+} signals and behaviours exhibited by growth cones. ■
 Timothy Gomez is in the Department of Anatomy, University of Wisconsin Medical School, 1300 University Avenue, Madison, Wisconsin 53706, USA. e-mail: tmgomez@wisc.edu

1. Wang, G. X. & Poo, M. *Nature* **434**, 898–904 (2005).
2. Li, Y. *et al.* *Nature* **434**, 894–898 (2005).
3. Hong, K., Nishiyama, M., Henley, J., Tessier-Lavigne, M. & Poo, M. *Nature* **403**, 93–98 (2000).

4. Montell, C., Birnbaumer, L. & Flockerzi, V. *Cell* **108**, 595–598 (2002).
5. Greka, A., Navarro, B., Oancea, E., Duggan, A. & Clapham, D. E. *Nature Neurosci.* **6**, 837–845 (2003).
6. Jin, M. *et al.* *J. Neurosci.* **25**, 2338–2347 (2005).
7. Wen, Z. X., Guirland, C., Ming, G. L. & Zheng, J. Q. *Neuron* **43**, 835–846 (2004).
8. Robles, E., Huttenlocher, A. & Gomez, T. M. *Neuron* **38**, 597–609 (2003).
9. Henley, J. R., Huang, K. H., Wang, D. & Poo, M. M. *Neuron* **44**, 909–916 (2004).
10. Nishiyama, M. *et al.* *Nature* **423**, 990–995 (2003).

Astrophysics

Two's company

Roger Cayrel

The matter from which the first stars formed was that left behind by the Big Bang. Stars containing extremely small amounts of heavy elements such as iron provide clues to the chemical composition of this matter.

Identifying stars born at the beginning of the era of stellar formation proved a long and frustrating task. In 2002, however, Christlieb *et al.*¹ reported the discovery of the first such 'relic of the dawn of time'. Writing on page 871 of this issue, Frebel *et al.*² announce the discovery of a second. But why are these two objects so remarkable? And why does it help to have two instead of one?

Some 13.7 billion years ago, the Universe was much simpler than it is today. It consisted of a uniform, hot gas showing only small fluctuations in temperature and density, and containing no large structures — no galaxies, stars or planets. For the first 15 minutes of its existence, the temperature and density of this hot gas were high enough to allow the nuclear reactions necessary for the production of the lightest chemical elements. Heavier elements, such as the metals, were not produced in this first flurry of nucleosynthetic activity. After the first 15 minutes, the Universe's rapid expansion put an end to conditions that favoured nucleosynthesis and nothing more happened to the nuclear composition of the Universe for about 200 million years. The ingredients of this frozen primordial soup — principally, the light elements lithium and helium, as well as deuterium, a heavy isotope of hydrogen — are fairly well known from both theory and experiment³.

A second opportunity for nuclear activity arose only when the original fluctuations of the early Universe had grown sufficiently large for haloes of dark matter to begin to form. This triggered gravitational instabilities and the collapse of conventional baryonic matter into clouds of gas, from which stars then formed⁴; in the cores of these stars, both the temperature and density reached values that again made nuclear reactions possible. Practically all the heavier elements, from carbon to uranium — the elements from which later solid planets and

organic life formed — were synthesized in these first stars.

The search for stars with a composition reflecting that of these first stars has been going on for the past 25 years^{5,6}. Before the discovery of the two 'relics' (HE0107–5240 by Christlieb *et al.*¹ and HE1327–2326 by Frebel *et al.*²), the lowest proportion, with respect to hydrogen, of stellar-made elements in the oldest stars was about a ten-thousandth of that observed in the Sun — a tiny amount, but, crucially, not zero. This finding seemed to support the theory that matter from the Big Bang was unable to fragment into stellar masses that were small enough for those stars still to be shining today: massive stars burn their fuel faster, and a star with a mass greater than around nine-tenths of the mass of the Sun would have exhausted its nuclear energy supply by now. Long-lived stars could thus be born only from interstellar gas already enriched in products of nucleosynthesis that had been expelled at the end of the evolution of earlier stars — either through a violent event such as a supernova, or through less dramatic mass loss, as is caused for example by stellar winds. If true, this would have ruled out any hope of finding a star with the primordial mix.

The discovery of HE0107–5240 and HE1327–2326, which have iron abundances respectively 200,000 and 300,000 times smaller than that of the Sun, is therefore of great significance. However, despite their impressively low iron content — well below the previous record, which stood for some 20 years before their discovery⁷ — both stars contain a proportion of carbon that is only 25 times smaller than that of the Sun. And the deficiency of the various elements between carbon and iron in the periodic table increases steadily with increasing atomic number.

The main difference between the two relics is that, in the case of HE0107–5240,

it proved impossible to determine whether lithium, carbon and nitrogen were still present in their original abundances. Frebel *et al.*², however, have been able to measure the abundance of lithium in HE1327–2326; surprisingly, it is far lower than that in other very old metal-poor stars. Other questions raised by HE1327–2326 are its abundance of nitrogen, which is 60 times larger than that of HE0107–5240; the shallower slope of the decrease of elemental abundances with atomic number; and the presence of strontium, not seen in HE0107–5240.

Previous work may help to unravel these mysteries. For example, it has been shown⁸ that a supernova may expel only the upper layers of the nuclear matter of the pre-supernova star, enriching the carbon and oxygen content of the interstellar matter, whereas the iron falls back on to the remnant of the star. The abundance of iron in a star is irrelevant to the fragmentation process⁹: carbon and oxygen are thought to be the elements controlling the cooling of the interstellar matter, and it is their high abundance that allows the star to fragment into small masses.

Rotation can also produce devastating effects on the evolution of massive stars with low metal abundances¹⁰: if its initial rotation is sufficiently great, the star experiences a significant loss of mass, shedding great quantities of the first elements synthesized — helium, carbon, oxygen and magnesium. It is exactly these elements that are among the most abundant in HE0107–5240 and HE1327–2326. Rotation also triggers an efficient mixing mechanism that allows the synthesis of nitrogen as well as carbon and oxygen. This is of significance for HE1327–2326, in which — after hydrogen and helium, both of primordial origin — nitrogen is the most abundant element.

It is fascinating to see that, having taken 20 years to undercut the previous lowest iron abundance in a star, we have within two years found two stars containing more than an order of magnitude less iron. This is the reward for an enormous effort in observational work that is bringing new insights to the fragmentation process that occurred during the formation of the first stars. ■

Roger Cayrel is at the Observatoire de Paris-Meudon, 61 Avenue de l'Observatoire, 75014 Paris, France. e-mail: roger.cayrel@obspm.fr

1. Christlieb, N. *et al.* *Nature* **419**, 904–906 (2002).
2. Frebel, A. *et al.* *Nature* **434**, 871–873 (2005).
3. Coc, A., Vangioni-Flam, E., Descouvemont, P., Adahchour, A. & Angulo, C. *Astrophys. J.* **600**, 544–552 (2004).
4. Bromm, V. & Larson, R. B. *Annu. Rev. Astron. Astrophys.* **42**, 79–118 (2004).
5. Beers, T. C., Preston, G. W. & Shectman, S. A. *Astron. J.* **103**, 1987–2034 (1992).
6. Christlieb, N. *et al.* in *ESO Astrophys. Symp.* (eds Weiss, A., Abel, T. G. & Hill, V.) 49–50 (Springer, Berlin, 2000).
7. Bessell, M. S. & Norris, J. *Astrophys. J.* **285**, 622–636 (1984).
8. Woosley, S. E. & Weaver, T. *Astrophys. Suppl. Ser.* **101**, 181–235 (1995).
9. Bromm, V. & Loeb, A. *Nature* **425**, 812–814 (2003).
10. Meynet, G., Maeder, A. & Ekström, S. preprint at www.arxiv.org/astro-ph/0408322 (2004).

DNA sequencing

Different dyes for clear-cut colours

Proc. Natl Acad. Sci. USA **102**, 5346–5351 (2005)

Since its introduction almost 20 years ago, four-colour DNA sequencing has largely relied on the same, somewhat error-prone, method. Now Ernest K. Lewis *et al.* have built a prototype sequencing machine that could improve accuracy.

In conventional colour sequencing, the chemical bases that make up DNA are tagged with fluorescent dyes — a different colour for each of the four bases. A machine shines a laser onto the DNA molecules, and detects the wavelength of light emitted from each base to determine their sequence. But mistakes happen, partly because the spectra produced by the dyes overlap, and hence the glow from one dye can be mistaken for that from another.

For the new method, called pulsed multi-line excitation, the researchers developed a different set of four fluorescent dyes, each of which is excited by a separate wavelength. Their machine fires a series of four laser beams at the dye, but only the appropriate laser triggers a signal. The method could greatly improve the ease with which one base can be distinguished from another.

Helen Pearson

Cancer

Remote control

Curr. Biol. **15**, 561–565 (2005)

BRCA1 is notorious as the first gene to be linked with inherited susceptibility to breast and ovarian cancer. It has been thought of as a classic ‘tumour suppressor’, but Rajas Chodankar *et al.* suggest that it may have another, more subtle, effect.

Granulosa cells in the ovary produce the sex hormones that regulate the ovulatory cycle — and the growth of ovarian tumours. Given that repeated ovulations (that is, fewer pregnancies or reduced oral contraceptive use) are known to increase the risk of non-hereditary ovarian cancer, the researchers wondered whether decreased levels of *BRCA1* protein in granulosa cells are involved. Using mice, they inactivated the gene specifically in these cells. The animals developed tumours in the ovaries and uterine horns. But the tumour cells looked like epithelial cells and had normal copies of the gene, implying that they had not developed from granulosa cells.

Inactivating *BRCA1* seems, therefore, to be controlling some intermediary produced by the granulosa cells. It is this unidentified factor that appears to promote tumours in the ovary epithelium, so providing a lead for further investigation.

Helen Dell



Particle physics

The elusive axion

Phys. Rev. Lett. **94**, 121301 (2005)

An effect known as charge–parity violation is linked to the fact that the Universe contains far more matter than antimatter, and it is well documented in processes involving the so-called weak nuclear force, one of the four fundamental forces of nature. But it seems to be suppressed by the strong force, and this can be explained by postulating a hitherto undiscovered particle, the axion. Axions interact hardly at all with radiation or other matter, making them hot candidates to be the ‘cold dark matter’ that is thought to pervade the Universe.

The CAST (CERN Axion Solar Telescope) collaboration has adopted an innovative approach to the search for axions. They are

pointing a powerful test magnet (pictured), decommissioned from CERN’s Large Hadron Collider, at the Sun. Axions might be produced in the solar plasma when photons are scattered in strong electromagnetic fields. CAST has put the scattering effect into reverse by producing X-ray photons from solar–axion interactions on Earth.

The magnet can be tilted at either end to an angle that allows the Sun to be observed at sunrise and sunset, both ends being fitted with X-ray detectors and an X-ray telescope recycled from the German space programme. The results, assuming a very small axion mass, show no signal above background, and constrain the axion–photon coupling strength by a factor of five compared with results from previous lab experiments. Future measurements should deliver still better sensitivity, and also test the axion hypothesis for higher masses.

Richard Webb

Neurobiology

Illuminating behaviour

Cell **121**, 141–152 (2005)

Through genetic engineering, researchers have developed a new technique for exciting neurons and influencing fruitfly behaviour. Whereas scientists typically excite these cells with electricity, the effect here was achieved with laser light.

Susana Q. Lima and Gero Miesenböck designed fruitflies to express particular ion channels in neurons that control escape mechanisms — such as jumping and wing beating — or in the dopamine-producing cells that influence movement. The next step involved injecting the flies with ATP (energy-storing molecules) held in chemical cages.

A 200-millisecond pulse of laser light — directed at the flies — removed the cage from the ATP molecules, allowing them to stimulate the channels and depolarize the neurons. When the authors targeted the neurons linked to escape mechanisms, the light set off jumping and wing flapping in the fruitflies. Similarly, targeting dopamine-producing cells altered the insects’ walking behaviour. The authors speculate that this ability to direct animal behaviour by remote control will enable them to study how specific behaviours are related to specific neurons.

Roxanne Khamsi

Spintronics

How electrons relax

Phys. Rev. Lett. **94**, 116601 (2005)

In the burgeoning field of spintronics, binary bits of data are stored in the spins of electrons, rather than in their charge, with a ‘1’ equating to spin up and a ‘0’ to spin down. But one problem facing the development of spintronic devices is that, although electron spin can be manipulated, it tends not to stay so — an induced spin decays as the electron interacts with the magnetic field of nearby nuclei.

P.-F. Braun and colleagues have now directly observed this ‘spin relaxation’ in quantum dots — clusters of atoms just nanometres across — made of the semiconductor materials indium arsenide and gallium arsenide. The authors found that the initial spin polarization of such dots decays with a half-life of just 0.5 nanoseconds — half a millionth of a millisecond — before remaining stable at about a third of its initial value for at least a further 10 nanoseconds.

However, they also report that this relaxation process can be suppressed by an externally applied static magnetic field of just 100 mT, which can be provided by small permanent magnets. Such a field increases the characteristic decay half-life to around 4 nanoseconds, and could prove useful in future practical devices, they suggest.

Mark Peplow

Low methane leakage from gas pipelines

A switch from coal or oil to natural gas could mitigate climate effects in the short term.

Using natural gas for fuel releases less carbon dioxide per unit of energy produced than burning oil or coal, but its production and transport are accompanied by emissions of methane, which is a much more potent greenhouse gas than carbon dioxide in the short term. This calls into question whether climate forcing could be reduced by switching from coal and oil to natural gas¹. We have made measurements in Russia along the world's largest gas-transport system and find that methane leakage is in the region of 1.4%, which is considerably less than expected and comparable to that from systems in the United States. Our calculations indicate that using natural gas in preference to other fossil fuels could be useful in the short term for mitigating climate change.

The global production of natural gas (which is about 90% methane) at present amounts to roughly 2,600 billion cubic metres (b.c.m.) a year². The high energy efficiency of modern gas-fired power plants (about 60%) and the large reserves of natural gas stimulate annual investments of about US\$100 billion (ref. 3). Russia, which has gas reserves of 47,000 b.c.m., is the largest producer (580 b.c.m. per year) and is the principal supplier of natural gas to the European Union (115 b.c.m. per year); the United States is the second-largest producer² (550 b.c.m. per year).

Methane has an atmospheric lifetime of about a decade, whereas carbon dioxide remains in the atmosphere–climate system for about a century. However, methane also makes an indirect chemical contribution to climate change, and its global-warming potential compared with carbon dioxide (mole/mole) is about 22 for a time horizon of 20 years, and about 8 for 100 years^{1,4}. This implies that a reduction in methane emissions could be beneficial in the mitigation of climate change over relatively short time-scales of a few decades or less.

For a short-term scenario (considering global-warming potential over 20 years), we estimate that switching from coal to natural gas would mitigate climate change if gas-associated leakage of methane was kept below $5.6 \pm 0.7\%$; if gas is swapped for oil, losses need to be less than $3.1 \pm 0.3\%$. For global-warming potential over 100 years, methane losses should remain under $11.3 \pm 0.7\%$ and $7.0 \pm 0.3\%$ for changing from coal and oil, respectively. We conclude that, provided methane leakage into the atmosphere is below about 3%, the use of natural gas as fuel should contribute least to climate forcing under all scenarios.

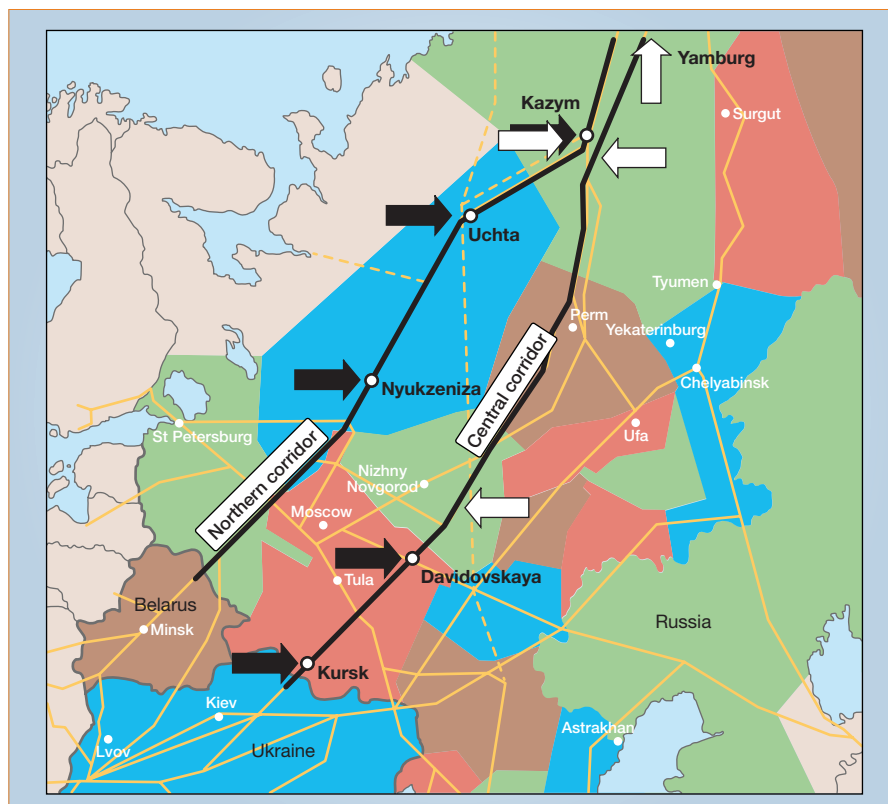


Figure 1 Russian pipelines used to transport natural gas. Orange lines, pipeline networks; dashed lines, pipelines under construction; black lines, pipelines that have been tested for methane leakage. Sites where measurements were taken in 1996/97 (ref. 6) are indicated by white arrows; black arrows indicate the sites where our measurements were made in 2003. (Coloured areas designate regional Gazprom divisions and different gas-transport companies in Russia, Ukraine and Belarus.)

It has been suggested that up to 10% of Russian gas is lost during production and transport⁵. Although some analyses indicate that leakages are lower than this^{5,6}, the database from which these figures were derived included only a few measurements. We therefore carried out extensive measurements of methane emissions along the Russian gas-transport system (Fig. 1), focusing on components from which leaks would be expected to occur and investigating nearly 2,400 km of pipelines, including compressor stations, valve knots and machine halls, as a representative selection of equipment types and ages (for details and methods, see supplementary information).

Our results indicate an overall loss of methane during natural-gas transport within Russia of 0.7% (range, 0.4–1.6%). Exporting gas may sometimes be associated with further release of methane, depending on transport conditions and distances (for example, the gas may be liquefied or pipeline extensions introduced). We assume that the onward distribution of gas in the Russian low-pressure networks for domestic and industrial purposes makes up 0.5–0.8% of

emissions, or about 30–50% of the total, as in the United States^{5,7,8}. To determine gas spills at wells, we re-evaluated earlier measurements⁶, arriving at a slightly higher leakage value of $0.1 \pm 0.04\%$. It is therefore likely that, overall, the leakage from Russian natural-gas transport systems is about 1.4% (with a range of 1.0–2.5%), which is comparable to the amount lost from pipelines in the United States ($1.5 \pm 0.5\%$)⁷.

Measurements from global air-sampling networks show that atmospheric methane increased by about 10% per decade between the 1960s and 1980s, but that this has levelled off since, owing in part to the reduced production of fossil fuels after the collapse of the Soviet Union in 1992 (refs 9, 10). This event also affected Russian natural-gas throughput, which varied by 7–8% over the past decade; however, worldwide gas production has increased by 25% since 1990 (refs 1, 2). The technology and maintenance of Russian gas-transfer systems may have improved⁵. It is likely that sources and sinks of atmospheric methane are now roughly balanced⁹ and that emission controls have taken effect even before implementation of the Kyoto Protocol.

We conclude that the use of natural gas as a preferred fuel, particularly as a replacement for coal, could reduce climate forcing during the transition to cleaner-energy technologies. Cutting methane emissions is cost effective and will be beneficial both directly and indirectly by reducing tropospheric ozone¹¹, even as atmospheric carbon dioxide continues to increase at an accelerating pace.

J. Lelieveld*, S. Lechtenböhmer†, S. S. Asanov*, C. A. M. Brenninkmeijer*, C. Dienst†, M. Fischedick†, T. Hanke†

*Max Planck Institute for Chemistry, 55128 Mainz, Germany

e-mail: lelieveld@mpch-mainz.mpg.de

†Wuppertal Institute for Climate, Environment and Energy, 42103 Wuppertal, Germany

1. Lelieveld, J. & Crutzen, P. J. *Nature* **355**, 339–342 (1992).

2. British Petroleum Statistical Review of World Energy 2004 <http://www.bp.com> (BP, London, 2004).

3. International Energy Agency Natural Gas Information 2004 (OECD/IEA, Paris, 2004).

4. Ramaswamy, V. et al. in *Climate Change 2001, the Third Assessment Report of the Intergovernmental Panel On Climate Change* (eds Houghton, J. T. et al.) 349–416 (Cambridge Univ. Press, Cambridge, UK, and New York, 2001).

5. Reshetnikov, A. I., Paramonova N. N. & Shashkov, A. A. *J. Geophys. Res.* **105**, 3517–3529 (2000).

6. Dedikov, J. V. et al. *Atmos. Environ.* **33**, 3291–3299 (1999).

7. US Environmental Protection Agency Inventory of US Greenhouse Gas Sources and Sinks: 1990–2002 <http://yosemite.epa.gov/oar/globalwarming.nsf> (2004).

8. US Environmental Protection Agency Methane Emissions from the Natural Gas Industry EPA-600/R-96-080 (1996).

9. Dlugokencky, E. J. et al. *Geophys. Res. Lett.* **30**, 1992 doi:10.1029/2003GL018126 (2003).

10. Wang, J. S. et al. *Glob. Biogeochem. Cycles* **18**, GB3011 doi:10.1029/2003GB002180 (2004).

11. Fiore, A. M. et al. *Geophys. Res. Lett.* **29**, 1919 doi:10.1029/2002GL015601 (2002).

Supplementary information accompanies this communication on Nature's website.

Competing financial interests: declared (see online brief communication).

Planetary science

Constant illumination at the lunar north pole

Images returned by the spacecraft Clementine have been used to produce a quantitative illumination map of the north pole of the Moon, revealing the percentage of time that points on the surface are illuminated during the lunar day. We have used this map to identify areas that are constantly illuminated during a lunar day in summer and which may therefore be in permanent sunlight. All are located on the northern rim of Peary crater, close to the north pole. Permanently sunlit areas represent prime locations for lunar outpost sites as they have abundant solar energy, are relatively benign thermally (when compared with equatorial regions), and are close to permanently shadowed regions that may contain water ice.

Because the Moon's spin axis is nearly perpendicular (about 1.5°) to the ecliptic plane, astronomers have long considered that areas of illumination extremes may exist near the lunar poles¹. Topographic lows,

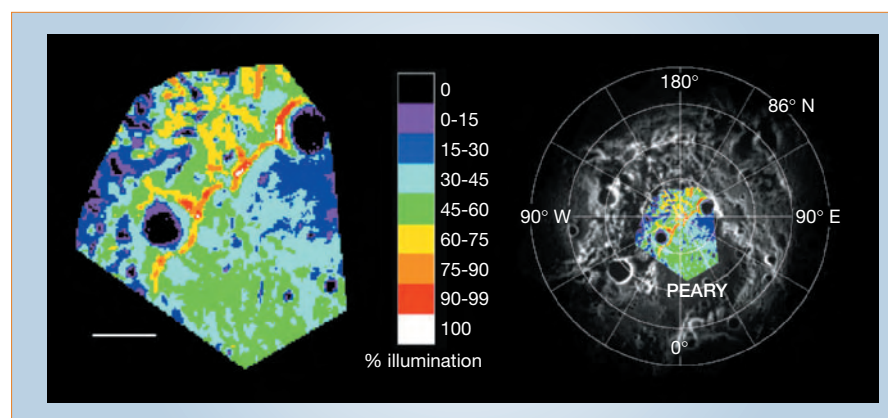


Figure 1 A quantitative illumination map of the Moon's north pole. The colour scale indicates the percentage of time that a point on the surface is illuminated during a lunar day in summer. Left, several areas on the rim of Peary crater (78 km in diameter) can be identified (white) that were continuously illuminated. The spatial extent of the map is within about 1–1.5° of the pole. Scale bar, 15 km. Right, colour illumination map superimposed on a greyscale composite Clementine mosaic, for spatial reference.

such as the floors of impact craters, can be permanently in shadow², whereas high areas could in theory be in constant sunlight. Permanently shadowed regions are extremely cold³ and may contain deposits of water ice^{4,5}.

Conversely, the temperature in a permanently lit zone is relatively mild and constant because of the grazing insolation. The temperature at the lunar equator fluctuates from –180°C to 100°C, but the surface temperature for a constantly sunlit polar region has been estimated from modelling⁶ to be roughly –50 ± 10°C. A region with this relatively benign temperature range represents an attractive site for building hardware designed for long-term use.

The lunar north pole is in a highland region, in between three large impact craters⁷ — Peary (88.6° N, 33.0° E; diameter, 73 km), Hermite (86.0° N, 89.9° W; diameter, 104 km) and Rozhdestvensky (85.2° N, 155.4° W; diameter, 177 km). Because the pole lies just outside the crater rims, it is likely to be at a relatively high elevation, increasing the likelihood that some areas could be permanently illuminated. By contrast, the lunar south pole is just inside the rim of the South Pole–Aitken impact basin (2,500-km diameter), and there is no area at this pole that is constantly illuminated during the southern winter, as measured at the scale of the Clementine UVVIS data (500 m per pixel)⁸.

The Clementine spacecraft orbited the Moon in an elliptical orbit with a 5-hour period for 71 days in 1994, arriving in the northern hemisphere just after mid-summer (that is, when the spin axis was pointing towards the Sun)⁹. We have identified 53 images taken by the craft's UVVIS camera that cover the north pole with a spatial resolution of roughly 500 m per pixel. Each image encompasses an area of about 190 × 140 km. The images show which areas are illuminated as a function of solar azimuth during a lunar day in summer.

Our quantitative illumination map for the north polar region shows the percentage

of time that a point on the surface is illuminated during a lunar summer day (Fig. 1). There are several regions, all on the rim of Peary crater, that are illuminated for the entire day (Fig. 1, white areas). With the information available, it is not possible to state definitively that these areas are permanently sunlit because the data correspond to a summer rather than a winter day. But we can be certain that they are the most illuminated regions around the north pole and that they are also the areas on the Moon most likely to be permanently sunlit, given that there are no constantly illuminated areas in the south polar region⁸.

Our quantitative illumination map also identifies permanently shadowed areas. These are associated with small impact craters (3 km or less in diameter) on the floor of Peary crater, with two larger craters (14 and 17 km in diameter) on the rim of Peary, and with the area just outside Peary's rim, in the highland region (Fig. 1).

D. Ben J. Bussey*, Kirsten E. Fristad†, Paul M. Schenk‡, Mark S. Robinson§, Paul D. Spudis*

*Planetary Exploration Group, Space Department, The Johns Hopkins Applied Physics Laboratory, Laurel, Maryland 20902, USA

e-mail: ben.bussey@jhuapl.edu

†Macalester College, Saint Paul, Minnesota 55105, USA

‡Lunar and Planetary Institute, Houston, Texas 77058, USA

§Center for Planetary Sciences, Northwestern University, Evanston, Illinois 60208, USA

1. Burke, J. D. in *Lunar Bases and Space Activities of the 21st Century* 77–84 (Lunar and Planetary Institute, Houston, 1985).

2. Bussey, D. B. J. et al. *Geophys. Res. Lett.* **30**, 1278–1281 (2003).

3. Vasavada, A. R., Paige, D. A. & Wood, S. E. *Icarus* **141**, 179–193 (1999).

4. Nozette, S. et al. *Science* **274**, 1495–1498 (1996).

5. Feldman, W. C. et al. *J. Geophys. Res.* **105**, 4175–4195 (2000).

6. Heiken, G., Vaniman, D. & French, B. M. (eds) *The Lunar Sourcebook* (Cambridge Univ. Press, Cambridge, UK, 1991).

7. Bussey, D. B. J. & Spudis, P. D. *The Clementine Atlas of the Moon* (Cambridge Univ. Press, Cambridge, UK, 2004).

8. Bussey, D. B. J., Spudis, P. D. & Robinson, M. S. *Geophys. Res. Lett.* **26**, 1187–1190 (1999).

9. Nozette, S. et al. *Science* **266**, 1835–1839 (1994).

Competing financial interests: declared none.

Wnt signalling in stem cells and cancer

Tannishtha Reya¹ & Hans Clevers²

¹Department of Pharmacology and Cancer Biology, Duke University Medical Center, Durham, North Carolina 27710, USA

²Hubrecht Laboratory, Uppsalalaan 8, 3584 CT Utrecht, The Netherlands

The canonical Wnt cascade has emerged as a critical regulator of stem cells. In many tissues, activation of Wnt signalling has also been associated with cancer. This has raised the possibility that the tightly regulated self-renewal mediated by Wnt signalling in stem and progenitor cells is subverted in cancer cells to allow malignant proliferation. Insights gained from understanding how the Wnt pathway is integrally involved in both stem cell and cancer cell maintenance and growth in the intestinal, epidermal and haematopoietic systems may serve as a paradigm for understanding the dual nature of self-renewal signals.

Stem cells are cells that have the unique ability to self-renew as well as to generate more differentiated progeny. The most primitive stem cell is the embryonic stem cell, which is derived from the inner cell mass of the blastocyst. This cell is pluripotent and can thus generate all the tissues of the body. Following the pioneering work on haematopoietic stem cells over the last five decades, a multitude of recent studies have indicated that most other adult tissues also harbour stem cells. These adult stem cells are normally involved in homeostatic self-renewal processes but can also be rapidly recruited to repair tissues upon injury. With the study of adult stem cell biology, recurring roles of a limited set of signalling cascades are rapidly being uncovered. One of these is the canonical Wnt cascade. Notably, in many of the same tissues where the Wnt cascade controls stem cells, cancer ensues upon dysregulated activation of this pathway. Conceptually, this indicates that an efficient road to cancer involves the hijacking of physiological regulators of stem cell function in these particular tissues. Below, we first outline the canonical Wnt cascade, and then describe its mirror image roles in the biology of stem cells and cancer.

The canonical Wnt signalling pathway in development

The discovery of the common origin of the *Drosophila* segment polarity gene *Wingless* and the murine proto-oncogene *Int-1* (ref. 1) laid the keystone of a signalling pathway now commonly referred to as the canonical Wnt cascade (Fig. 1). Wnt genes, of which the human genome harbours almost 20, occur throughout the animal kingdom. Signalling is initiated when Wnt ligands engage their cognate receptor complex, consisting of a serpentine receptor of the Frizzled family and a member of the LDL receptor family, Lrp5/6. The central player is a cytoplasmic protein termed β -catenin, the stability of which is regulated by the destruction complex. There are β -catenin-independent, non-canonical signalling pathways induced by Wnt, and their mechanism of action has been described elsewhere². When Wnt receptors are not engaged, two scaffolding proteins in the destruction complex—the tumour suppressors adenomatous polyposis coli (APC) and axin—bind newly synthesized β -catenin. CKI and GSK3, two kinases residing in the destruction complex, then sequentially phosphorylate a set of conserved Ser and Thr residues in the amino terminus of β -catenin. The resulting phosphorylated footprint recruits a β -TrCP-containing E3 ubiquitin ligase, which targets β -catenin for proteasomal degradation.

Receptor occupancy inhibits the kinase activity of the destruction complex by an incompletely understood mechanism involving the direct interaction of axin with Lrp5/6, and/or the actions of an axin-binding molecule, Dishevelled. As a consequence, β -catenin accumulates and travels into the nucleus where it engages the N terminus of DNA-binding proteins of the Tcf/Lef family³. The

vertebrate genome encodes four highly similar Tcf/Lef proteins. In the absence of a Wnt signal, Tcf/Lef proteins repress target genes through a direct association with co-repressors such as Groucho. The interaction with β -catenin transiently converts Tcf/Lef factors into transcriptional activators. *Drosophila* genetics has recently identified two additional nuclear components, Pygopus and Bcl9 (also known as legless), conserved in vertebrates. Pygopus is essential for transcriptional activation of Tcf/Lef target genes, whereas Bcl9 seems to bridge Pygopus to Tcf-bound β -catenin. In sum, the canonical pathway translates a Wnt signal into the transient transcription of a Tcf/Lef target gene programme⁴.

From crypt physiology to colon cancer

The intestinal tract consists of the small intestine (duodenum,

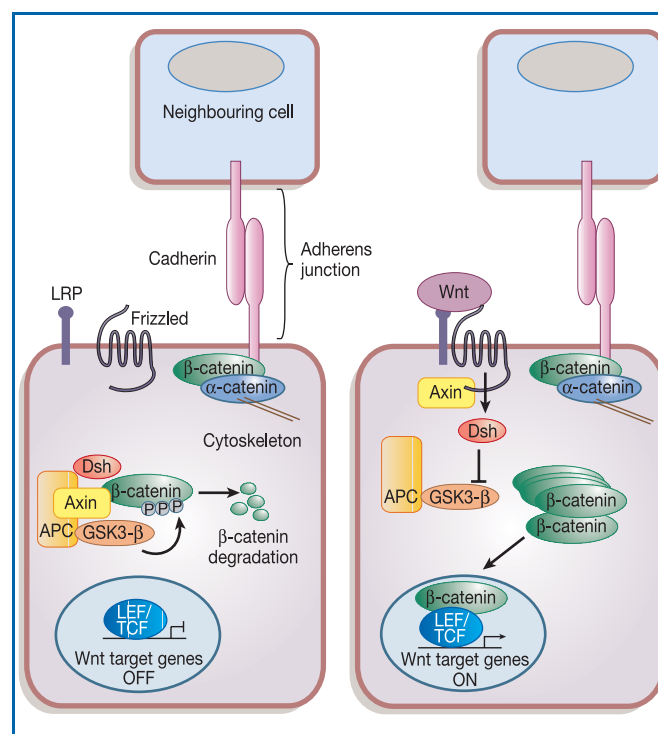


Figure 1 The canonical Wnt signalling pathway. In the absence of Wnt signalling (left panel), β -catenin is in a complex with axin, APC and GSK3- β , and gets phosphorylated and targeted for degradation. β -Catenin also exists in a cadherin-bound form and regulates cell–cell adhesion. In the presence of Wnt signalling (right panel), β -catenin is uncoupled from the degradation complex and translocates to the nucleus, where it binds Lef/Tcf transcription factors, thus activating target genes. (Adapted from ref. 44.)

jejunum and ileum) and the large intestine or colon. The absorptive epithelium of the small intestine is ordered into villi and crypts of Lieberkühn. Differentiated cells (enterocytes, enteroendocrine cells and goblet cells) occupy the villi (Fig. 2). A fourth differentiated type, the Paneth cell, resides at the bottom of crypts and secretes antimicrobial agents⁵. The remainder of the crypts constitutes the stem/progenitor compartment. The mucosa of the colon has a flat surface epithelium instead of villi (Fig. 3). Proliferative stem and precursor cells occupy the bottom two-thirds of crypts, whereas differentiated cells constitute the surface epithelium and top third of the crypts.

The life cycle of an individual epithelial cell spans less than a week. In the mouse, some 200 cells are generated per crypt every day. This is compensated by cell loss at the villus tip. The epithelium forms a contiguous two-dimensional sheet that is in a perpetual upward movement⁶. Stem cells and Paneth cells at the crypt bottom escape this flow⁷. Although no unique markers have been identified for intestinal stem cells, they can be efficiently labelled *in vivo* with 3H-thymidine or 5-bromodeoxyuridine (BrdU) during early postnatal life or after irradiation⁸. Long-term label retention,

mouse chimaeras^{9,10} and the study of regeneration upon injury have allowed an operational definition of numbers and localization. The estimated number of stem cells is between one and six per crypt⁷. Colon stem cells reside at the crypt bottom, whereas small intestinal stem cells are at 'position +4' above the Paneth cells (reviewed in ref. 11). These stem cells cycle slowly, continuously producing rapidly proliferating 'transit amplifying' cells capable of differentiating towards all lineages. They not only self-renew throughout life but also regenerate the epithelium after injury. Committed progenitors undergo cell cycle arrest and start expressing differentiation markers when they reach the top one-third of colorectal crypts or the crypt–villus junction in the small intestine. The differentiated cells continue their migration on the villus in coherent bands stretching along the crypt–villus axis. Although crypts are monoclonal, each villus receives cells from multiple crypts and is therefore polyclonal^{10,12}.

Current evidence indicates that the Wnt cascade is the single most dominant force in controlling cell fate along the crypt–villus axis. Nuclear β -catenin, the hallmark of Wnt signalling, is observed throughout the crypts of the intestine¹³. In *Tcf4*^{-/-} neonatal mice, the villus epithelial compartment appears unaffected but the crypt progenitor compartment is entirely absent, implying that physiological Wnt signalling is required for maintenance of the crypt progenitor phenotype¹⁴ (Table 1). In concordance, inhibition of the Wnt receptor complex by transgenic Dickkopf-1 expression induces the complete loss of crypts in adult mice^{15,16}. As mentioned below, a constitutive β -catenin–Tcf4 complex drives a genetic programme in colorectal cancer cells that is physiologically expressed in crypt stem/progenitor cells¹⁷. The Wnt signalling gradient also controls expression of the EphB/ephrinB sorting receptors and ligands¹³. The resulting EphB/ephrinB counter-gradients allow the establishment of crypt–villus boundaries as well as the positioning of Paneth cells at the crypt bottom.

The Wnt pathway in colon cancer

The *APC* gene was originally discovered to be the culprit in a hereditary cancer syndrome termed familial adenomatous polyposis (FAP). FAP patients, inheriting one defective *APC* allele, develop large numbers of colon polyps, or adenomas, early in life. Individual polyps are clonal outgrowths of epithelial cells in which the second *APC* allele is inactivated. The large burden of FAP adenomas inevitably results in the appearance of adenocarcinomas through clonal evolution, evident as a more or less ordered accumulation of mutations in additional oncogenes or tumour suppressor genes, such as *K-Ras*, *p53* and *Smad4*. Loss of *APC* also occurs in most sporadic colorectal cancers¹⁸. Mutational inactivation of *APC* leads to the inappropriate stabilization of β -catenin¹⁹, implying that the absence of functional *APC* transforms epithelial cells through activation of the Wnt cascade. Indeed, reporter plasmids containing concatemeric Tcf binding sites such as pTOPFLASH, normally transcribed only upon Wnt signalling, are inappropriately transcribed in *APC* mutant cancer cells through the action of constitutive β -catenin–Tcf4 complexes²⁰. In some cases of colorectal cancer in which *APC* is not mutated, the scaffolding protein axin 2 is mutant²¹, or activating (oncogenic) point mutations in β -catenin remove its N-terminal Ser/Thr destruction motif²². As exceptions to the rule that *APC* inactivation initiates colon cancer, these rare mutations underscore the central role of the inappropriate persistence of β -catenin–Tcf4 complexes in the transformation of epithelial cells.

Multiple studies have used a candidate gene approach to address the nature of the Tcf4 target gene programme in colorectal cancer. Among others, c-Myc²³ and cyclin D1 (ref. 24) reportedly are essential components of this programme in transformed intestinal epithelial cells. We recently induced expression of dominant-negative Tcf4 in colorectal cell lines and carried out a global assessment of the target gene programme using DNA microarrays¹⁷.

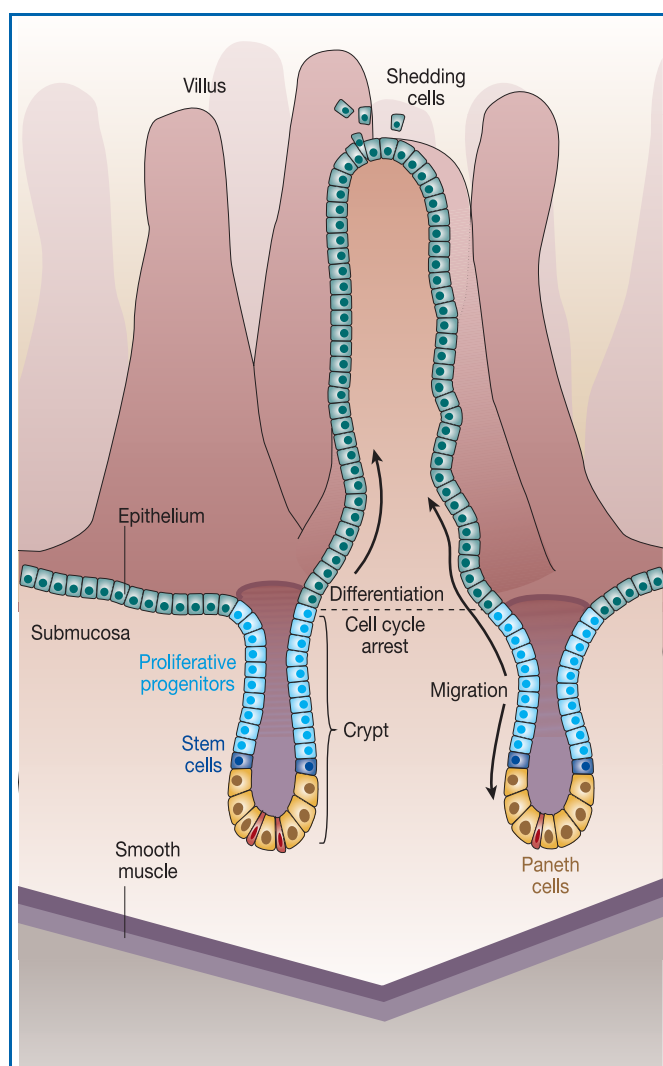


Figure 2 Tissue anatomy of the adult small intestine. Putative stem cells (dark blue) reside immediately above the Paneth cells (yellow) near the crypt bottom. Proliferating progenitor cells occupy the remainder of the crypt. Differentiated cells (green) populate the villus, and include goblet cells, enterocytes and enteroendocrine cells. (Adapted from ref. 89.)

This study revealed that Tcf4 drives the same genetic programme in colorectal cancer cells as in crypt stem and progenitor cells. *APC*^{-/-} adenoma cells thus represent the transformed counterparts of the proliferative crypt progenitor. Once the Wnt cascade is mutationally activated, the adenoma cells maintain their progenitor status indefinitely. This allows the adenomas to persist for many years, providing ample opportunity for the acquisition of further mutations. In an elegant *in vivo* study, the Wnt cascade was mutationally activated by inducible deletion of an *APC* allele. Within 1–2 days, villi were entirely populated by crypt-like cells. DNA array analysis revealed the induction of many of the crypt markers previously identified as Tcf4 targets in human colorectal cancer cell lines²⁵.

Colorectal cancer almost invariably initiates with an activating mutation in the Wnt cascade. Current evidence indicates that the transformation process exploits the unique physiological dependence of the intestinal progenitor/stem cell on the Wnt cascade. A surprising symmetry ensues between the crypt progenitor and adenoma cells.

Wnt signalling in epidermal development and cancer

The hair follicle and the associated sebaceous gland are appendices of the squamous interfollicular epidermis of the skin²⁶. Hair follicles

are far more complex structures than the crypt–villus unit described above. Multipotent epidermal stem cells reside in the bulge region, a structure situated below the sebaceous gland at the site of attachment of the arrector pili muscle (Fig. 4). Bulge stem cells can generate the hair lineages, sebocytes, as well as the stem cells of the interfollicular epidermis. Distinct stem cell populations may exist also in the interfollicular epidermis and sebaceous gland (see ref. 26). To form a hair, cells flow downward from the bulge through the outer root sheath. At the base of the follicle, the cells enter a transit amplifying compartment termed the germinative matrix. After exiting the matrix, the cells undergo terminal differentiation in the precortex compartment. Bulge-derived cells can also flow upward through the outer root sheath to yield sebocytes, or to seed the basal layer of the interfollicular epidermis with slow-cycling keratinocyte stem cells. These latter cells maintain a transit amplifying compartment, in which cells divide several times before the onset of terminal keratinocyte differentiation.

Mice with a mutation in one of the four *Tcf/Lef* genes, *Lef1*, display a paucity of hair follicles²⁷ (Table 1). Conversely, overexpression of transgenic *Lef1* in bulge and interfollicular epidermis leads to hair germ-like invaginations in the epidermis²⁸. A much more dramatic phenotype is obtained upon overexpression of a constitutively stable form of β -catenin²⁹, which induces pro-

Table 1 Summary of selected studies demonstrating the effects of Wnt signalling on stem cells from varied tissues

Stimulus	Cell type	Species	Effect	Assay	Reference
Epidermal stem and progenitor cells					
Activated β -catenin	Keratinocytes	Mouse	Inhibition of differentiation	<i>In vitro</i>	83
Activated β -catenin transgenic	Follicular stem cells	Mouse	Hair follicle and tumour formation	<i>In vivo</i>	29
<i>Lef1</i> ^{-/-}	Hair follicles	Mouse	Loss of hair formation	<i>In vivo</i>	27
<i>Lef1</i> transgenic	Hair follicles	Mouse	Increased hair follicle formation	<i>In vivo</i>	28
Conditional deletion of β -catenin	Follicular stem cells	Mouse	Cell fate alteration	<i>In vivo</i>	31
Intestinal stem and progenitor cells					
<i>Tcf4</i> ^{-/-}	Intestinal stem cells	Mouse	Depletion of crypt stem cells	<i>In vivo</i>	20
<i>Dkk1</i> transgenic	Epithelial cells in intestine	Mouse	Loss of crypts; reduced epithelial proliferation	<i>In vivo</i>	15
Adenoviral <i>Dkk1</i>	Epithelial cells in intestine	Mouse	Reduced epithelial proliferation	<i>In vivo</i>	16
Haematopoietic stem and progenitor cells					
<i>Tcf1</i> ^{-/-}	T-cell progenitors	Mouse	Loss of early proliferative progenitors	<i>In vivo</i>	84
<i>Lef1</i> ^{-/-} <i>Tcf1</i> ^{-/-}	T-cell progenitors	Mouse	Loss of early proliferative progenitors; block in T-cell differentiation	<i>In vivo</i>	85
<i>Wnt1</i> ^{-/-} <i>Wnt4</i> ^{-/-}	T-cell progenitors	Mouse	Reduced thymic proliferation and cellularity	<i>In vivo</i>	86
Axin inducible transgenic	T-cell progenitors	Mouse	Reduced thymic cellularity and increased apoptosis	<i>In vivo</i>	87
Conditional deletion of β -catenin (lck-driven)	T-cell progenitors	Mouse	Block in differentiation, reduced proliferation	<i>In vivo</i>	51
<i>Lef1</i> ^{-/-}	B-cell progenitors	Mouse	Reduced survival and proliferation	<i>In vivo</i>	52
<i>Fzd9</i> ^{-/-}	B-cell progenitors	Mouse	Loss of early B-cell progenitors	<i>In vivo</i>	62
Wnt5A + SLF	Fetal liver enriched precursors	Mouse	Increased progenitors	<i>In vitro</i>	46
Wnt5A or Wnt11A	Bone marrow, non-adherent	Quail	Increased differentiation	<i>In vitro</i>	82
Wnt3A + stromal cells	Bone marrow, unenriched	Mouse	Inhibition of haematopoiesis	<i>In vitro</i>	88
Wnt3A/activated β -catenin	Purified HSCs	Mouse	Increased self renewal	<i>In vitro/in vivo</i>	45, 47
Axin	Purified HSCs	Mouse	Inhibition of proliferation/reconstitution	<i>In vitro/in vivo</i>	45
Conditional deletion of β -catenin	Bone marrow	Mouse	Normal self renewal/ reconstitution	<i>In vivo</i>	50
Wnt5A + SLF + feeder cells	Bone marrow enriched precursors	Human	Increased progenitors	<i>In vitro</i>	48
Wnt5A	Bone marrow enriched precursors	Human	Increased reconstitution	<i>In vivo</i> (xenotransplant)	49
Neural stem and progenitors					
Activated β -catenin transgenic	Neural progenitors	Mouse	Increased expansion	<i>In vivo</i>	67
Conditional deletion of β -catenin	Neural progenitors	Mouse	Decreased expansion	<i>In vivo</i>	68
Conditional activation of β -catenin	Neural progenitors	Mouse	Increased expansion	<i>In vivo</i>	68
Embryonic stem cells					
GSK3-specific inhibitor	Embryonic stem cells	Mouse/human	Maintenance of pluripotency	<i>In vitro</i>	74
6-bromoindirubin-3'-oxime (BIO)	Embryonic stem cells	Mouse	Increased inhibition of differentiation	<i>In vivo/in vitro</i>	75
APC mutations	Embryonic stem cells	Mouse	Increased inhibition of differentiation	<i>In vivo/in vitro</i>	75

in the gut, pilomatricomas may have hijacked the Wnt-driven expansion of hair precursors in the matrix/precortex region of the hair. Whereas several aspects of hair follicle biology are thus controlled by the Wnt cascade, it remains undetermined whether the Wnt pathway is also required in the interfollicular epidermis, or is mutated in tumours derived thereof.

Wnt signalling in haematopoietic stem cells

Haematopoietic stem cells (HSCs), the cells that give rise to the blood throughout our lifetime, are perhaps the best understood stem cells in our body. These cells were originally identified functionally by Till and McCulloch^{37,38}. Since then, great progress has been made in understanding the biology of HSCs, in large part because they have been successfully isolated and the stages through which they progress during development delineated (reviewed in ref. 39). HSC activity in the adult bone marrow has been found to lie within a very rare population of cells characterized by their expression of undetectable or low levels of haematopoietic lineage markers, high levels of c-kit and Sea-1 and low levels of Thy-1.1³⁹. These cells probably reside in the context of a stromal cell niche, which, at least in part, consists of cells of the osteoblast lineage (Fig. 5)^{57,58}. The ability to isolate HSCs and progenitors has also allowed recent advancement in the understanding of the molecular control of their function, particularly that of self-renewal and maintenance. There is growing evidence that signals such as Notch, Hedgehog and Wnt—pathways that control many developmental processes and are dysregulated in cancer—might regulate self-renewal of haematopoietic progenitors and stem cells^{40–43}. In the case of Wnt signalling, the ligands (that is, Wnt proteins) are produced by HSCs themselves as well as by the microenvironment, suggesting that HSCs may use them in an autocrine or paracrine manner (reviewed in ref. 44). Furthermore, the observation of Wnt reporter activity in HSCs in the context of their native

microenvironment suggests that HSCs directly respond to Wnt signalling *in vivo*⁴⁵. Functional studies show that *in vitro*, soluble Wnt proteins synergize with Steel factor (also known as stem cell factor or SLF) to promote the growth and inhibit the differentiation of murine haematopoietic progenitors⁴⁶ (Table 1). In addition, β -catenin as well as purified Wnt3A protein can promote self renewal of murine HSCs *in vitro* and enhance their ability to reconstitute the haematopoietic system of lethally irradiated mice *in vivo*^{45,47}. The effects of Wnt signalling on HSCs in mice are recapitulated in human HSCs and progenitors as well. Wnt5A treatment of human haematopoietic progenitors in the presence of stromal cell contact promotes the expansion of undifferentiated progenitors *in vitro*⁴⁸, and treatment of mice with Wnt5A-conditioned medium results in increased human HSC re-population in a NOD-SCID xenotransplant model⁴⁹. Cumulatively, these studies suggest that Wnt signalling can contribute to HSC and progenitor cell self-renewal. Because Wnt proteins that act through the canonical pathway (such as Wnt3A) and those that can act through non-canonical pathways (for example, Wnt5A) seem to influence HSCs and progenitors equivalently, the pathway dominantly used during HSC function *in vivo* remains a question for further study.

That Wnt signalling may also be required for normal growth of HSCs *in vitro* and *in vivo* is suggested by inhibition of HSC growth and decreased reconstitution after ectopic expression of axin⁴⁵; however, mice in which β -catenin has been conditionally deleted revealed no haematolymphoid defects, raising questions about the role of β -catenin in HSCs and in haematopoietic progenitor development⁵⁰. These data are difficult to reconcile with the previous literature that suggests that β -catenin and other elements of the Wnt signalling pathway are required for T-cell development⁵¹, B-cell survival^{52,53}, and haematopoietic stem and progenitor function (reviewed in ref. 54). One possibility is that the interferon-induced deletion of β -catenin sets up a significantly different context for study of the requirement of specific genes in HSC development⁵¹; alternatively, compensation by γ -catenin, a close relative of β -catenin, may allow β -catenin mutant HSCs to function normally⁵⁵. This is supported by the recent finding that γ -catenin overexpression leads to enhanced re-plating efficiency of HSCs, selectively causing expansion of small, uncommitted, blast-like colonies *in vitro*⁵⁶, and the fact that axin can enhance degradation not only of β -catenin but of γ -catenin as well. Analyses of mice mutant for other components of the Wnt pathway are clearly needed to elucidate the precise nature of its influence on HSC maintenance and self renewal *in vivo*. Understanding the influence of Wnt signalling on HSCs in the context of the native microenvironment will also be enhanced by identifying the function of the specific Wnt proteins produced at the HSC niche^{57,58}. Because the niche is also likely to provide other signals that have been shown to influence HSCs (such as Shh and Notch ligands⁵⁹), it will be critical to determine the roles that these signals have relative to Wnts and whether they control distinct elements of self renewal.

Aberrant Wnt signalling in leukaemia

Much like the parallels between adenomas and progenitor cells in the intestinal crypt, the influence of Wnt pathway components on haematopoietic stem and progenitor cells suggests that the same pathway may be dysregulated to cause leukaemia. Although this possibility has been speculated on for some time, it has only recently been supported by experimental evidence demonstrating that oncogenic growth in leukaemias of both myeloid and lymphoid lineages is dependent on Wnt signalling.

Myeloid leukaemias can be broadly classified as chronic or acute, and Wnt signalling has been implicated in regulating the growth of cells from both subtypes of leukaemia. Chronic myelogenous leukaemia results predominantly from a BCR-Abl translocation that originates in HSCs but exhibits its consequences in more committed myeloid precursors. Granulocyte-macrophage progeni-

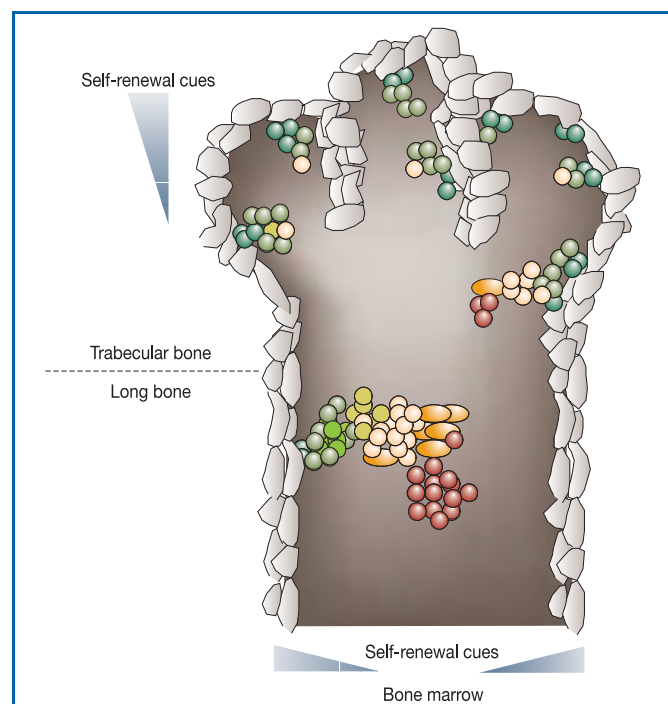


Figure 5 Proposed model of HSC development in the niche. HSCs are shown (dark green) at the endosteal marrow adjacent to the bone's surface mainly at the trabecular bone, and are postulated to migrate inward in the central marrow as they differentiate (precursors in light green; differentiated cells in yellow, orange and red) away from a possible gradient of self-renewal cues. (Adapted from ref. 44.)

tors (GMPs) from chronic myelogenous leukaemia patients and blast crisis cells from patients resistant to therapy display activated Wnt signalling, as determined by Wnt reporter activity and accumulation of nuclear β -catenin⁶⁰. Additionally, inhibition of β -catenin through ectopic expression of axin decreases the replating capacity of leukaemic cells *in vitro*, suggesting that chronic myelogenous leukaemia precursors are dependent on Wnt signalling for growth and renewal. In an interesting parallel with its effects on normal HSCs, overactivation of Wnt signalling also endowed GMPs—which normally have limited self-renewal capacity—with stem-cell-like properties of long-term renewal, at least *in vitro*⁶⁰. This supports the notion that for progenitor cells to be transformed they must adopt a stem-cell-like ability to renew, and suggests that Wnt signalling is a cue that can control such a transformation event. The possibility that leukaemic cells depend on Wnt signalling is supported by studies of acute myelogenous leukaemia, which show that the re-plating efficiency of HSCs transduced with fusion proteins encoded by translocations commonly found in acute myelogenous leukaemia was abrogated upon inhibition of γ -catenin^{56,61}. Conversely, mice that were transplanted with HSCs overexpressing γ -catenin displayed acute myelogenous leukaemia-like symptoms⁵⁶. Although these emerging data suggest that Wnt activation through β - and γ -catenin may lead to myeloid leukaemias, drawing a causative link awaits generation of a mouse model where aberrant Wnt signalling can drive leukaemogenesis.

Recent studies suggest that lymphoid neoplasias may also be influenced by Wnt signalling. Wnt16 was initially cloned owing to its overexpression in pre-B-cell leukaemia lines carrying the E2A-PbX translocation⁶², raising the possibility that autocrine Wnt usage may contribute to oncogenesis. This possibility is supported by the observed influence of Wnt signalling on growth and survival of normal B-cell progenitors^{52,53}. A similar autocrine loop of Wnt dependence has also been proposed for regulating the growth of multiple myeloma, a cancer of terminally differentiated B cells⁶³. A

number of primary myelomas and myeloma cell lines were found to express the stabilized form of β -catenin; although no mutations in Wnt signalling components were detected, highly increased levels of Wnt proteins including Wnt5A and Wnt10B were detected. In an interesting parallel with the influence of the Shh pathway in cancers of the lung and gastrointestinal tract⁶⁴, these data suggest that leukaemias may not harbour the type of Wnt pathway mutations found in colon cancer or epidermal tumours, but they may nonetheless be critically dependent on the autocrine/paracrine use of Wnt signalling for sustaining cancerous self-renewal. Paradoxically, although in most contexts Wnt proteins induce proliferation and are associated with oncogenesis, a small subset of Wnt5A hemizygous mice develop lymphoid and myeloid leukaemia⁶⁵. This suggests that in some contexts a Wnt5A-mediated non-canonical pathway inhibits the canonical pathway, perhaps at the level of Tcf phosphorylation and function⁶⁶, and that the loss of this suppression may lead to hyperproliferation of B-cell precursors and subsequent leukaemogenesis. Thus, antagonism between the canonical pathway and non-canonical pathways may be an essential mechanism for controlling dysregulated renewal and oncogenesis.

Perspectives

The concept that Wnt signalling may preferentially influence progenitor cell expansion in development and in cancer extends beyond the examples discussed above. Recent data show that activating Wnt signalling through β -catenin can increase cycling and expansion of neural progenitors, and that its loss causes a reduction in the progenitor compartment^{67,68}. Just as normal activation of Wnt signalling may promote self-renewal of neuronal stem cells, aberrant Wnt pathway activation may be tumorigenic in the nervous system. This is supported by the fact that medulloblastoma, a paediatric brain tumour of the cerebellum, harbours mutations in β -catenin⁶⁹ and axin^{70,71}, suggesting that some medulloblastomas may arise from a primitive progenitor that

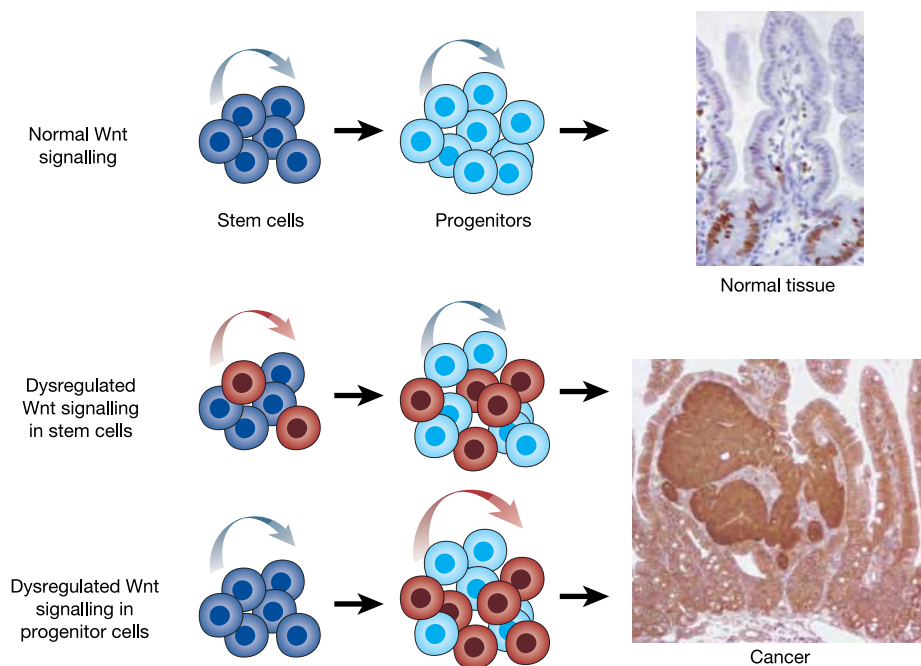


Figure 6 Normal Wnt signalling influences the proliferation and renewal of stem cells (dark blue) or progenitors (light blue) during development of a variety of tissues. Uncontrolled Wnt signalling (red arrow) can lead to constitutive renewal and aberrant expansion of the stem cell pool, or confer stem cell behaviour (long-term

renewal) on the progenitor cell pool. These changes can, in turn, lead to formation of cancerous tissues. The cancerous cells are denoted in red in the schematic diagram.

becomes transformed in response to uncontrolled Wnt signalling. Furthermore, in mammary tissues where stem cells have yet to be definitively isolated, a controlling role for Wnt in progenitor cell fate or maintenance is suggested by studies of Wnt1 transgenic mice that develop mammary tumours. These tumours have an increased frequency of cells with stem and progenitor properties^{72,73}, in contrast to tumours from mice overexpressing other oncogenes. This suggests again that the Wnt pathway may be unique in its ability to target stem and progenitor cells for transformation, and perhaps reflects a role of the Wnt pathway in self-renewal of normal breast epithelium⁷³. Finally, the influence of Wnt signalling on stem cells appears to extend to the most primitive of all stem cells: the embryonic stem cell. Modulation of Wnt signalling by controlling the dose of APC, or by treatment with a GSK3- β inhibitor, can enhance self-renewal of both mouse and human embryonic stem cells^{74,75}.

Looking to the future, it is clear that an important area of investigation will involve analysing how the Wnt pathway exerts its effect and whether shared molecular targets are involved in influencing self-renewal in the context of stem cells and cancer¹⁷. Additionally, Wnt signalling probably integrates with other niche-derived signals such as BMP, Shh and Notch^{76–78}. In some contexts, these signals may be independently responsible for distinct aspects of tissue self-renewal, such as survival, proliferation and inhibition of differentiation. In other cases, the various signalling cascades may act in a hierarchy, and regulate each other. It will be critical to understand the coordinated activity of these pathways in the *in vivo* control of stem cell function and define whether the same patterns hold true in cancer. Finally, although much of the focus of this review is on Wnt-controlled cell renewal, there is no doubt that, depending on the context, Wnt signalling can influence commitment and differentiation as well^{79–82}. A clearer view of the different cellular and molecular contexts that elicit differential outcomes of Wnt signalling will be essential to understand more completely the basis of Wnt's effects on renewal in stem cells and cancer.

This review highlights how the Wnt pathway may be a common element in regulating stem cell/progenitor cell renewal and maintenance in a variety of systems, and how this ability is exploited by cancer (Fig. 6). Because Wnt signalling can act to maintain stem cells as well as cancer cells, the ability to modulate the Wnt pathway either positively or negatively may be of therapeutic relevance. Thus, controlled activation of Wnt signalling may allow us to enhance stem cell and progenitor cell activity, when regeneration is needed. On the other hand, inhibition of Wnt signalling may prove to be an effective road to inhibiting the uncontrolled renewal that drives cancer. □

doi:10.1038/nature03319.

- Rijsewijk, F. *et al.* The *Drosophila* homolog of the mouse mammary oncogene int-1 is identical to the segment polarity gene wingless. *Cell* **50**, 649–657 (1987).
- Veeman, M. T., Axelrod, J. D. & Moon, R. T. A second canon. Functions and mechanisms of β -catenin-independent Wnt signaling. *Dev. Cell* **5**, 367–377 (2003).
- Eastman, Q. & Grosschedl, R. Regulation of LEF-1/TCF transcription factors by Wnt and other signals. *Curr. Opin. Cell Biol.* **11**, 233–240 (1999).
- Giles, R. H., van Es, J. H. & Clevers, H. Caught up in a Wnt storm: Wnt signaling in cancer. *Biochim. Biophys. Acta* **1653**, 1–24 (2003).
- Porter, E. M., Bevins, C. L., Ghosh, D. & Ganz, T. The multifaceted Paneth cell. *Cell. Mol. Life Sci.* **59**, 156–170 (2002).
- Heath, J. P. Epithelial cell migration in the intestine. *Cell Biol. Int.* **20**, 139–146 (1996).
- Potten, C. S. Stem cells in gastrointestinal epithelium: numbers, characteristics and death. *Phil. Trans. R. Soc. Lond. B* **353**, 821–830 (1998).
- Potten, C. S., Owen, G. & Booth, D. Intestinal stem cells protect their genome by selective segregation of template DNA strands. *J. Cell Sci.* **115**, 2381–2388 (2002).
- Bjerknes, M. & Cheng, H. Clonal analysis of mouse intestinal epithelial progenitors. *Gastroenterology* **116**, 7–14 (1999).
- Schmidt, G. H., Winton, D. J. & Ponder, B. A. Development of the pattern of cell renewal in the crypt-villus unit of chimaeric mouse small intestine. *Development* **103**, 785–790 (1988).
- Booth, C. & Potten, C. S. Gut instincts: thoughts on intestinal epithelial stem cells. *J. Clin. Invest.* **105**, 1493–1499 (2000).
- Roth, K. A., Hermiston, M. L. & Gordon, J. I. Use of transgenic mice to infer the biological properties of small intestinal stem cells and to examine the lineage relationships of their descendants. *Proc. Natl Acad. Sci. USA* **88**, 9407–9411 (1991).

13. Batle, E. *et al.* β -catenin and TCF mediate cell positioning in the intestinal epithelium by controlling the expression of EphB/ephrinB. *Cell* **111**, 251–263 (2002).
14. Korinek, V. *et al.* Depletion of epithelial stem-cell compartments in the small intestine of mice lacking Tcf-4. *Nature Genet.* **19**, 1–5 (1998).
15. Pinto, D., Gregorieff, A., Begthel, H. & Clevers, H. Canonical Wnt signals are essential for homeostasis of the intestinal epithelium. *Genes Dev.* **17**, 1709–1713 (2003).
16. Kuhnert, F. *et al.* Essential requirement for Wnt signaling in proliferation of adult small intestine and colon revealed by adenoviral expression of Dickkopf-1. *Proc. Natl Acad. Sci. USA* **101**, 266–271 (2004).
17. van de Wetering, M. *et al.* The β -catenin/TCF-4 complex imposes a crypt progenitor phenotype on colorectal cancer cells. *Cell* **111**, 241–250 (2002).
18. Kinzler, K. W. & Vogelstein, B. Lessons from hereditary colorectal cancer. *Cell* **87**, 159–170 (1996).
19. Rubinfeld, B. *et al.* Binding of GSK3 β to the APC- β -catenin complex and regulation of complex assembly. *Science* **272**, 1023–1026 (1996).
20. Korinek, V. *et al.* Constitutive transcriptional activation by a β -catenin-Tcf complex in APC-/- colon carcinoma. *Science* **275**, 1784–1787 (1997).
21. Liu, W. *et al.* Mutations in AXIN2 cause colorectal cancer with defective mismatch repair by activating β -catenin/TCF signalling. *Nature Genet.* **26**, 146–147 (2000).
22. Morin, P. J. *et al.* Activation of β -catenin-Tcf signaling in colon cancer by mutations in β -catenin or APC. *Science* **275**, 1787–1790 (1997).
23. He, T. C. *et al.* Identification of c-MYC as a target of the APC pathway. *Science* **281**, 1509–1512 (1998).
24. Tetsu, O. & McCormick, F. β -catenin regulates expression of cyclin D1 in colon carcinoma cells. *Nature* **398**, 422–426 (1999).
25. Sansom, O. J. *et al.* Loss of Apc *in vivo* immediately perturbs Wnt signaling, differentiation, and migration. *Genes Dev.* **18**, 1385–1390 (2004).
26. Niemann, C. & Watt, F. M. Designer skin: lineage commitment in postnatal epidermis. *Trends Cell Biol.* **12**, 185–192 (2002).
27. van Genderen, C. *et al.* Development of several organs that require inductive epithelial-mesenchymal interactions is impaired in LEF-1 deficient mice. *Genes Dev.* **8**, 2691–2703 (1994).
28. Zhou, P., Byrne, C., Jacobs, J. & Fuchs, E. Lymphoid enhancer factor 1 directs hair follicle patterning and epithelial cell fate. *Genes Dev.* **9**, 700–713 (1995).
29. Gat, U., DasGupta, R., Degenstein, L. & Fuchs, E. *De Novo* hair follicle morphogenesis and hair tumors in mice expressing a truncated β -catenin in skin. *Cell* **95**, 605–614 (1998).
30. Lo Celso, C., Prowse, D. M. & Watt, F. M. Transient activation of β -catenin signalling in adult mouse epidermis is sufficient to induce new hair follicles but continuous activation is required to maintain hair follicle tumours. *Development* **131**, 1787–1799 (2004).
31. Hulsken, J., Vogel, R., Erdmann, B., Cotsarelis, G. & Birchmeier, W. β -Catenin controls hair follicle morphogenesis and stem cell differentiation in the skin. *Cell* **105**, 533–545 (2001).
32. DasGupta, R. & Fuchs, E. Multiple roles for activated LEF/TCF transcription complexes during hair follicle development and differentiation. *Development* **126**, 4557–4568 (1999).
33. Niemann, C., Owens, D. M., Hulsken, J., Birchmeier, W. & Watt, F. M. Expression of Δ NLEF1 in mouse epidermis results in differentiation of hair follicles into squamous epidermal cysts and formation of skin tumours. *Development* **129**, 95–109 (2002).
34. Braun, K. M. *et al.* Manipulation of stem cell proliferation and lineage commitment: visualisation of label-retaining cells in whole mounts of mouse epidermis. *Development* **130**, 5241–5255 (2003).
35. Merrill, B. J., Gat, U., DasGupta, R. & Fuchs, E. Tcf3 and Leftf1 regulate lineage differentiation of multipotent stem cells in skin. *Genes Dev.* **15**, 1688–1705 (2001).
36. Chan, E. F., Gat, U., McNiff, J. M. & Fuchs, E. A common human skin tumour is caused by activating mutations in β -catenin. *Nature Genet.* **21**, 410–413 (1999).
37. Till, J. & McCulloch, E. A direct measurement of the radiation sensitivity of normal mouse bone marrow cells. *Radiat. Res.* **14**, 213–222 (1961).
38. Becker, A. J., McCulloch, E. A. & Till, J. E. Cytological demonstration of the clonal nature of spleen colonies derived from transplanted mouse marrow cells. *Nature* **197**, 452–454 (1963).
39. Weissman, I. L. Translating stem and progenitor cell biology to the clinic: barriers and opportunities. *Science* **287**, 1442–1446 (2000).
40. Taipale, J. & Beachy, P. A. The Hedgehog and Wnt signalling pathways in cancer. *Nature* **411**, 349–354 (2001).
41. Reya, T., Morrison, S. J., Clarke, M. F. & Weissman, I. L. Stem cells, cancer, and cancer stem cells. *Nature* **414**, 105–111 (2001).
42. Varnum-Finney, B. *et al.* Pluripotent, cytokine-dependent, hematopoietic stem cells are immortalized by constitutive Notch1 signaling. *Nature Med.* **6**, 1278–1281 (2000).
43. Bhardwaj, G. *et al.* Sonic hedgehog induces the proliferation of primitive human hematopoietic cells via BMP regulation. *Nature Immunol.* **2**, 172–180 (2001).
44. Rattis, F. M., Voermans, C. & Reya, T. Wnt signaling in the stem cell niche. *Curr. Opin. Hematol.* **11**, 88–94 (2004).
45. Reya, T. *et al.* A role for Wnt signalling in self-renewal of haematopoietic stem cells. *Nature* **423**, 409–414 (2003).
46. Austin, T. W., Solar, G. P., Ziegler, F. C., Liem, L. & Matthews, W. A role for the Wnt gene family in hematopoiesis: expansion of multilineage progenitor cells. *Blood* **89**, 3624–3635 (1997).
47. Willert, K. *et al.* Wnt proteins are lipid-modified and can act as stem cell growth factors. *Nature* **423**, 448–452 (2003).
48. Van Den Berg, D. J., Sharma, A. K., Bruno, E. & Hoffman, R. Role of members of the Wnt gene family in human hematopoiesis. *Blood* **92**, 3189–3202 (1998).
49. Murdoch, B. *et al.* Wnt-5A augments repopulating capacity and primitive hematopoietic development of human blood stem cells *in vivo*. *Proc. Natl Acad. Sci. USA* **100**, 3422–3427 (2003).
50. Cobas, M. *et al.* β -Catenin is dispensable for hematopoiesis and lymphopoiesis. *J. Exp. Med.* **199**, 221–229 (2004).
51. Xu, Y., Banerjee, D., Hulsken, J., Birchmeier, W. & Sen, J. M. Deletion of β -catenin impairs T cell development. *Nature Immunol.* **4**, 1177–1182 (2003).
52. Reya, T. *et al.* Wnt signaling regulates B lymphocyte proliferation through a LEF-1 dependent mechanism. *Immunity* **13**, 15–24 (2000).
53. Ranheim, E., Kwan, H., Reya, T., Weissman, I. L. & Francke, U. Frizzled 9 knockout mice have abnormal B cell development. *Blood* ahead of print publication, 30 November 2004 (doi:10.1182/blood-2004-06-2334).
54. van de Wetering, M., de Lau, W. & Clevers, H. WNT signaling and lymphocyte development. *Cell* **109**(suppl.), S13–S19 (2002).

55. Cowin, P., Kapprell, H. P. & Franke, W. W. The complement of desmosomal plaque proteins in different cell types. *J. Cell Biol.* **101**, 1442–1454 (1985).
56. Zheng, X. *et al.* γ -catenin contributes to leukemogenesis induced by AML-associated translocation products by increasing the self-renewal of very primitive progenitor cells. *Blood* **103**, 3535–3543 (2004).
57. Calvi, L. M. *et al.* Osteoblastic cells regulate the haematopoietic stem cell niche. *Nature* **425**, 841–846 (2003).
58. Zhang, J. *et al.* Identification of the haematopoietic stem cell niche and control of the niche size. *Nature* **425**, 836–841 (2003).
59. Hackney, J. A. *et al.* A molecular profile of a hematopoietic stem cell niche. *Proc. Natl Acad. Sci. USA* **99**, 13061–13066 (2002).
60. Jamieson, C. H. *et al.* Granulocyte/macrophage progenitors in chronic myelogenous leukemia are candidate leukemia stem cells that activate the β -catenin pathway. *N. Engl. J. Med.* **351**, 657–667 (2004).
61. Muller-Tidow, C. *et al.* Translocation products in acute myeloid leukemia activate the Wnt signaling pathway in hematopoietic cells. *Mol. Cell. Biol.* **24**, 2890–2904 (2004).
62. McWhirter, J. R. *et al.* Oncogenic homeodomain transcription factor E2A-Pbx1 activates a novel WNT gene in pre-B acute lymphoblastoid leukemia. *Proc. Natl Acad. Sci. USA* **96**, 11464–11469 (1999).
63. Derksen, P. W. *et al.* Illegitimate WNT signaling promotes proliferation of multiple myeloma cells. *Proc. Natl Acad. Sci. USA* **101**, 6122–6127 (2004).
64. Watkins, D. N. *et al.* Hedgehog signalling within airway epithelial progenitors and in small-cell lung cancer. *Nature* **422**, 313–317 (2003).
65. Liang, H. *et al.* Wnt5a inhibits B cell proliferation and functions as a tumor suppressor in hematopoietic tissue. *Cancer Cell* **4**, 349–360 (2003).
66. Smit, L. *et al.* Wnt activates the Tak1/Nemo-like kinase pathway. *J. Biol. Chem.* **279**, 17232–17240 (2004).
67. Chenn, A. & Walsh, C. A. Regulation of cerebral cortical size by control of cell cycle exit in neural precursors. *Science* **297**, 365–369 (2002).
68. Zechner, D. *et al.* β -catenin signals regulate cell growth and the balance between progenitor cell expansion and differentiation in the nervous system. *Dev. Biol.* **258**, 406–418 (2003).
69. Zurawel, R. H., Chiappa, S. A., Allen, C. & Raffel, C. Sporadic medulloblastomas contain oncogenic β -catenin mutations. *Cancer Res.* **58**, 896–899 (1998).
70. Dahmen, R. P. *et al.* Deletions of AXIN1, a component of the WNT/wingless pathway, in sporadic medulloblastomas. *Cancer Res.* **61**, 7039–7043 (2001).
71. Baeza, N., Masuoka, J., Kleihues, P. & Ohgaki, H. AXIN1 mutations but not deletions in cerebellar medulloblastomas. *Oncogene* **22**, 632–636 (2003).
72. Liu, B. Y., McDermott, S. P., Khwaja, S. S. & Alexander, C. M. The transforming activity of Wnt effectors correlates with their ability to induce the accumulation of mammary progenitor cells. *Proc. Natl Acad. Sci. USA* **101**, 4158–4163 (2004).
73. Li, Y. *et al.* Evidence that transgenes encoding components of the Wnt signaling pathway preferentially induce mammary cancers from progenitor cells. *Proc. Natl Acad. Sci. USA* **100**, 15853–15858 (2003).
74. Sato, N., Meijer, L., Skaltsounis, L., Greengard, P. & Brivanlou, A. H. Maintenance of pluripotency in human and mouse embryonic stem cells through activation of Wnt signaling by a pharmacological GSK-3-specific inhibitor. *Nature Med.* **10**, 55–63 (2004).
75. Kielman, M. F. *et al.* Apc modulates embryonic stem-cell differentiation by controlling the dosage of β -catenin signaling. *Nature Genet.* **32**, 594–605 (2002).
76. Duncan, A. W. *et al.* Integration of Notch and Wnt signaling in hematopoietic stem cell maintenance. *Nature Immunol.* **6**(3), 314–322 (2005).
77. Galceran, J., Sustmann, C., Hsu, S. C., Folberth, S. & Grosschedl, R. LEF1-mediated regulation of Delta-like1 links Wnt and Notch signaling in somitogenesis. *Genes Dev.* **18**, 2718–2723 (2004).
78. He, X. C. *et al.* BMP signaling inhibits intestinal stem cell self-renewal through suppression of Wnt- β -catenin signalling. *Nature Genet.* **36**, 1117–1121 (2004).
79. Viti, J., Gulacsi, A. & Lillien, L. Wnt regulation of progenitor maturation in the cortex depends on Shh or fibroblast growth factor 2. *J. Neurosci.* **23**, 5919–5927 (2003).
80. Lee, H. Y. *et al.* Instructive role of Wnt/ β -catenin in sensory fate specification in neural crest stem cells. *Science* **303**, 1020–1023 (2004).
81. Muroyama, Y., Kondoh, H. & Takada, S. Wnt proteins promote neuronal differentiation in neural stem cell culture. *Biochem. Biophys. Res. Commun.* **313**, 915–921 (2004).
82. Brandon, C., Eisenberg, L. M. & Eisenberg, C. A. WNT signaling modulates the diversification of hematopoietic cells. *Blood* **96**, 4132–4141 (2000).
83. Zhu, A. J. & Watt, F. M. β -catenin signalling modulates proliferative potential of human epidermal keratinocytes independently of intercellular adhesion. *Development* **126**, 2285–2298 (1999).
84. Verbeek, S. *et al.* An HMG-box-containing T-cell factor required for thymocyte differentiation. *Nature* **374**, 70–74 (1995).
85. Okamura, R. M., Sigvardsson, M., Verbeek, S., Clevers, H. & Grosschedl, R. Redundant regulation of T cell differentiation and TCR α gene expression by the transcription factors LEF-1 and TCF-1. *Immunity* **8**, 11–20 (1998).
86. Mulroy, T., McMahon, J. A., Burakoff, S. J., McMahon, A. P. & Sen, J. Wnt-1 and Wnt-4 regulate thymic cellularity. *Eur. J. Immunol.* **32**, 967–971 (2002).
87. Hsu, W., Shakya, R. & Costantini, F. Impaired mammary gland and lymphoid development caused by inducible expression of Axin in transgenic mice. *J. Cell Biol.* **155**, 1055–1064 (2001).
88. Yamane, T. *et al.* Wnt signaling regulates hemopoiesis through stromal cells. *J. Immunol.* **167**, 765–772 (2001).
89. Sancho, E., Batlle, E. & Clevers, H. Signaling pathways in intestinal development and cancer. *Annu. Rev. Cell Dev. Biol.* **20**, 695–723 (2004).
90. Turksen, K. Revisiting the bulge. *Dev. Cell.* **6**, 454–456 (2004).

Acknowledgements We would like to thank F. Watt and I. Weissman for comments and suggestions, and F. Rattis for help with figures. T.R. is supported by an NIH grant and investigator awards from the Cancer Research Foundation and Ellison Medical Foundation. H.C. is supported by the Center for Biomedical Genetics, Cancer Genomics Consortium, SPINOZA, the Louis Jeantet-Foundation and the Dutch Cancer Foundation KWF.

Competing interests statement The authors declare that they have no competing financial interests.

Correspondence and requests for materials should be addressed to T.R. (t.reya@duke.edu) or H.C. (clevers@niob.knaw.nl).

Lead isotopes reveal bilateral asymmetry and vertical continuity in the Hawaiian mantle plume

W. Abouchami¹, A. W. Hofmann¹, S. J. G. Galer¹, F. A. Frey², J. Eisele¹ & M. Feigenson³

¹Max-Planck-Institut für Chemie, Postfach 3060, 55020 Mainz, Germany

²Massachusetts Institute of Technology, Cambridge, Massachusetts 02139, USA

³Department of Geological Sciences, Rutgers University, New Brunswick, New Jersey 08903, USA

The two parallel chains of Hawaiian volcanoes ('Loa' and 'Kea') are known to have statistically different but overlapping radiogenic isotope characteristics. This has been explained by a model of a concentrically zoned mantle plume, where the Kea chain preferentially samples a more peripheral portion of the plume. Using high-precision lead isotope data for both centrally and peripherally located volcanoes, we show here that the two trends have very little compositional overlap and instead reveal bilateral, non-concentric plume zones, probably derived from the plume source in the mantle. On a smaller scale, along the Kea chain, there are isotopic differences between the youngest lavas from the Mauna Kea and Kilauea volcanoes, but the 550-thousand-year-old Mauna Kea lavas are isotopically identical to Kilauea lavas, consistent with Mauna Kea's position relative to the plume, which was then similar to that of present-day Kilauea. We therefore conclude that narrow (less than 50 kilometres wide) compositional streaks, as well as the larger-scale bilateral zonation, are vertically continuous over tens to hundreds of kilometres within the plume.

Much of the worldwide isotopic heterogeneity in basalts from ocean island volcanoes, thought to be derived from mantle plumes, can be found in basalts derived from a single plume. Such is the case for the basalts from the Hawaiian islands. Recent Hawaiian volcanoes form a double chain widely referred to as the Loa and Kea trends, named after their largest volcanoes — Mauna Loa and Mauna Kea^{1–3}. It has long been known that systematic differences in Pb, Sr and Nd isotopic compositions exist between these two chains, but the published isotope data show large areas of overlap (see GEOROC (Geochemistry of Rocks of the Oceans and Continents) database <http://georoc.mpch-mainz.gwdg.de> and references therein), and the significance of these two trends remains unclear^{4,5}. One of the objectives of the Hawaii Scientific Drilling Project (HSDP) was to test the idea that the plume's isotopic and chemical composition is concentrically zoned as a result of the entrainment of external mantle material during plume ascent^{6–9}. Such entrainment might help to explain the compositional differences between volcanoes from the Kea and Loa chains. Phase 2 of HSDP drilling reached a depth of ~3 km and samples approximately 0.3 million years (Myr) of plume history¹⁰, corresponding to about 40 km of the horizontal traverse of Mauna Kea across the plume.

Recent HSDP-2 results have led to contrasting views of the geometry of the compositional heterogeneities across the plume conduit. Eisele *et al.*¹¹ interpreted the three distinct Pb isotope arrays (labelled Kea-lo8, Kea-mid8 and Kea-hi8) as reflecting vertical streaks at least several tens of kilometres in length within the Hawaiian plume stem. This interpretation implies that the volcano samples successive vertical streaks within the plume as it is carried across the plume by the Pacific plate. The alternative model¹² suggests that these heterogeneities are distributed vertically as a stack of 'pancakes', which are sampled consecutively by the volcano stratigraphy. Both vertical and horizontal variability are likely, but dynamic modelling suggests a dominance of vertical stretching during plume ascent¹³.

To address both questions—that of the larger-scale, possibly concentric plume zonation expressed by the Loa and Kea chains, and the significance of the smaller-scale variability observed within

a single Kea volcano (HSDP)—we report here high-precision Pb isotope ratios from nine recent Hawaiian volcanoes belonging to the Loa (Loihi, Mauna Loa, Lanai, Kahoolawe, Koolau) and Kea (Kilauea, Mauna Kea, Kohala) chains using a Pb triple-spike technique¹⁴. The sampling of these volcanoes includes early, pre-shield stage (Loihi), main-phase shield building (Kilauea, Kohala, Mauna Kea, Mauna Loa, Lanai, Kahoolawe, Koolau) and late-stage, tholeiitic to alkaline lavas (Mauna Kea) (Supplementary Table 1).

On the basis of these data, we demonstrate geographic and temporal isotopic variations in several Hawaiian volcanoes which indicate that the heterogeneities within the plume resemble vertically drawn-out 'streaks' rather than flat pancakes. We further show that the Hawaiian plume is compositionally not purely concentric. Rather, there exists a lateral chemical asymmetry, as expressed in the Loa–Kea chains. We discuss two scales of plume heterogeneity, beginning with local variations within the Kea chain.

Small-scale heterogeneity in the plume

Our data have a bearing on the origin of short-term heterogeneities observed within the stratigraphic section of Mauna Kea^{11,15}, which reveals small-scale features of the Hawaiian plume anatomy. In particular, the Pb isotopic composition of the submarine section of Mauna Kea is nearly identical to that of present-day Kilauea. Figure 1a shows a Pb isotopic comparison of different portions of the HSDP-1¹⁵ and HSDP-2¹¹ drill cores in Mauna Kea with those of recent lavas from Kilauea (red circles)—including Mauna Ulu and Hilina lavas—as well as Loihi (black triangles).

Among Mauna Kea lavas, there is a temporal decrease in ²⁰⁸Pb/²⁰⁴Pb at a given ²⁰⁶Pb/²⁰⁴Pb ratio (Fig. 1a). That is, the youngest (<350 kyr) Mauna Kea lavas, the uppermost 800 m of the HSDP drill core (green diamonds and squares) and subaerially exposed postshield lavas (green crosses) have the lowest ²⁰⁸Pb/²⁰⁴Pb, whereas older (350–550 kyr) Mauna Kea lavas, 800–3,000 m within the HSDP core, have higher ²⁰⁸Pb/²⁰⁴Pb (red and yellow diamonds)¹¹. The upper part of the HSDP core (green symbols) contains Pb isotope signatures that are quite distinct from those of recent Kilauea lavas (red circles). By contrast, the lower HSDP

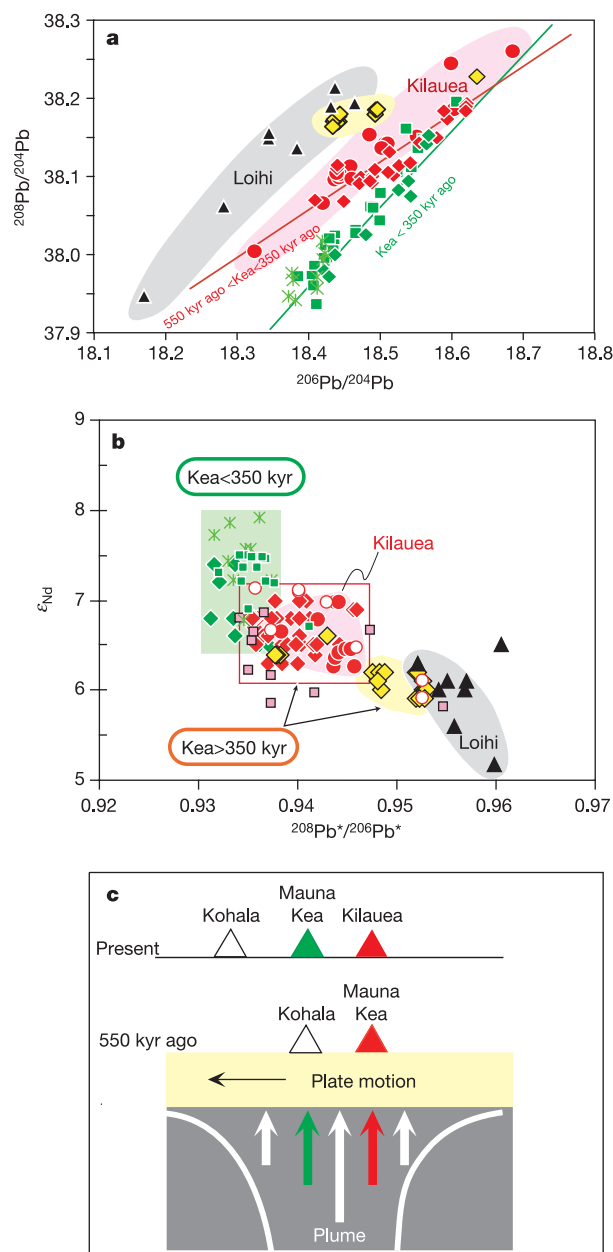


Figure 1 Small-scale heterogeneity in the Hawaiian plume revealed by the HSDP core, Mauna Kea. **a**, $^{208}\text{Pb}/^{204}\text{Pb}$ versus $^{206}\text{Pb}/^{204}\text{Pb}$ in HSDP-2 (ref. 11) (diamonds) and HSDP-1 lavas¹⁵ (green squares). Also plotted are Kilauea (red circles), Loihi (black triangles) and Mauna Kea post-shield lavas (green crosses). The HSDP-2 data are subdivided following Eisele *et al.*¹¹ into ‘Kea-hi8’ (yellow diamonds), ‘Kea-mid8’ (red diamonds) corresponding to samples with ages 350–550 kyr, and ‘Kea-lo8’ (green diamonds) with ages younger than 350 kyr. Most of the 350–550-kyr-old lavas fall in the field of Kilauea, with the exception of the Kea-hi8 samples, which have Pb isotopic compositions trending towards those of Loihi. Error bars on Pb isotope ratios are similar to the symbol sizes (2s.d. = 100 p.p.m.). **b**, ϵ_{Nd} versus $^{208}\text{Pb}^*/^{206}\text{Pb}^*$ ratios in HSDP-2 lavas. Same symbols as in **a**. Pb isotope data are from refs 11 and 15, and Nd isotope data from ref. 18, except for the red open circles, which correspond to HSDP-2 Nd data from this work. Comparison between HSDP-2 and Kilauea data obtained in our laboratory confirm the similarities between 350–550-kyr-old Kea samples and present-day Kilauea lavas. Like Loihi lavas, the Kea-hi8 samples (yellow diamonds) display high $^{208}\text{Pb}/^{204}\text{Pb}$ ratios and low ϵ_{Nd} . Pb and Nd isotope data on submarine Mauna Kea glasses (pink squares) are from ref. 21. **c**, Schematic illustration of the position of Mauna Kea relative to the plume 550 kyr ago, that is, the same location above the plume occupied by Kilauea today. The red arrow illustrates how similar plume sources are sampled by the two volcanoes at different times.

core (red diamonds) mostly comprises Kea-mid8 compositions¹¹ that are indistinguishable in Pb isotope space from young Kilauea lavas. The only exceptions are the apparently anomalous, Loihi-like lavas from very restricted stratigraphic levels in the vicinity of 1,800–2,100 m and 2,200–2,500 m depths in the HSDP core (yellow diamonds, Fig. 1a). These lavas share the low SiO_2 content that is typical of Loihi lavas¹⁶, as well as other Loihi characteristics, such as lower ϵ_{Nd} values and high $^3\text{He}/^4\text{He}$ ratios¹⁷. These exceptional compositions are enigmatic at present, and we can only speculate that they represent a local heterogeneity within the plume resembling Loihi sources in several respects. The recurrence of these lavas as intercalated flows in the deeper sections of the core suggests that this heterogeneity must have been a short-term feature of magma production during the main shield-building stage of Mauna Kea. We emphasize that, except for these anomalous lavas, the remainder of the 800–3,000 m section of Mauna Kea, with estimated eruption ages of 350–550 kyr, are indistinguishable from recent Kilauea lavas in $^{207}\text{Pb}/^{204}\text{Pb}$, $^{208}\text{Pb}/^{204}\text{Pb}$ – $^{206}\text{Pb}/^{204}\text{Pb}$ space.

Strontium and Nd isotope data do not provide the same level of volcano-specific resolution. Strontium isotope ratios in Mauna Kea, Kilauea and Loihi lavas are almost indistinguishable, ranging from 0.7034 to 0.7036 (see the GEOROC database). In contrast, published Nd isotope data and those reported here (Supplementary Table 1) indicate that Kilauea lavas have, on average, lower ϵ_{Nd} values (+6 to +7) than those measured in HSDP-1 (ref. 8) (mean = 7.19 ± 0.57 , $n=26$). HSDP-2 Mauna Kea lavas show a range of ϵ_{Nd} between 7.4 and 5.9 with an average of 6.6 ± 0.69 ($n=96$)¹⁸, lower than that reported for HSDP-1 (ref. 8). We also measured Nd isotopic compositions in seven HSDP-2 samples to compare these directly with Nd data on Kilauea lavas obtained in our laboratory. In $^{208}\text{Pb}^*/^{206}\text{Pb}^*$ – ϵ_{Nd} space (Fig. 1b)—where $^{208}\text{Pb}^*/^{206}\text{Pb}^*$ represents the time-integrated $^{232}\text{Th}/^{238}\text{U}$ ratio since the formation of the Earth and is defined as: $^{208}\text{Pb}^*/^{206}\text{Pb}^* = (^{208}\text{Pb}/^{204}\text{Pb} - 29.475)/(^{206}\text{Pb}/^{204}\text{Pb} - 9.306)$ (ref. 19)—the 350–550-kyr-old samples also fall in the field defined by Kilauea lavas. Thus, Nd isotope data, including our own, provide additional evidence for the similarities in sources sampled by 350–550-kyr-old Mauna Kea and present-day Kilauea. Submarine Mauna Kea glasses have also been shown to have radiogenic isotopic characteristics and trace-element ratios similar to those of Kilauea and Loihi lavas^{20,21} (Fig. 1b), and they are consistent with these observations.

The possibility that the 350–550-kyr-old samples are actually intercalated Kilauea lava flows can be ruled out, because the lowermost HSDP-2 samples are much older than the oldest-known Kilauea lavas (Hilina section dated 200–300 kyr ago (ref. 22); East Rift Zone, 350 kyr ago) (ref. 23). Moreover, recent data indicate that volcanism at Kilauea was in the early alkalic phase 220–275 kyr ago (ref. 24). The most straightforward explanation for the isotopic similarities is that the Kilauea-like lavas found in the HSDP-2 core are derived from Kilauea-type sources within the Hawaiian plume. The overall resemblance between young Kilauea and old Mauna Kea lavas can therefore be reasonably interpreted to result from the position of Mauna Kea relative to the upwelling plume centre. Reconstruction of the palaeo-distance of Mauna Kea from a hypothetical plume centre^{11,25} indicates a movement of the volcano by about 40 km during the past 550 kyr (from ~20 km 550 kyr ago to ~60 km at present), implying that Mauna Kea occupied almost the same relative spot then as Kilauea does now. This reconstruction provides a natural explanation for the isotopic similarities shared by the two volcanoes (Fig. 1a, c), and implies that the surface expression of the Hawaiian plume reflects precisely the position of the volcano relative to the plume axis at a given evolutionary stage of the volcano.

If old Mauna Kea sources are identical to recent Kilauea sources, we can use this information to estimate the vertical extent of this particular isotopic fingerprint within the plume, assuming we know something about ascent rates in the plume conduit. These rates have been estimated by a variety of methods^{11,26} that suggest an extreme

minimum given by the velocity of the Pacific plate of about 10 cm yr^{-1} and a conservative maximum of about 1 myr^{-1} near the centre of the plume^{27–29}. Using these upwelling velocities, the vertical distance travelled by plume material over $\sim 0.5 \text{ Myr}$ would be 50–500 km, implying that the minimum vertical extent of the specific ‘Kilauea-like’ signature common to Kilauea and Mauna Kea lies somewhere between 50 and 500 km. These considerations suggest that the Hawaiian plume contains relatively narrow compositional streaks (‘spaghetti’) that are drawn out by laminar flow in the conduit over a significant part of the vertical extent of the plume¹³. The sampling of different plume streaks by a volcano as it moves across the plume is illustrated in Fig. 1c. Because the characteristic isotopic signature of each streak is a linear array in Pb isotope space, each streak can be envisioned as consisting of two isotopically distinct endmember components, mixing in variable proportions during melting to produce a linear array^{11,15}. The fluctuations along these linear arrays reflect an additional, still smaller scale of isotopic variations, corresponding to a dominant frequency of about 10,000 yr (ref. 11) and a spatial scale of 5–50 km, depending on the plume ascent rate, but this feature is beyond the scope of this paper. The occurrence of several Pb isotope arrays during the lifetime of Mauna Kea reflects sampling of isotopically distinct streaks as the volcano moves laterally over the plume owing to plate motion (Supplementary Fig. S2).

Large-scale heterogeneity of the plume

A much larger-scale heterogeneity of the Hawaiian plume is revealed by the isotope data shown in Fig. 2a. The striking systematic differences between Loa (blue symbols) and Kea (red symbols) volcanoes demonstrate a clear left-right compositional division within the Hawaiian plume. In $^{208}\text{Pb}/^{204}\text{Pb}$ – $^{206}\text{Pb}/^{204}\text{Pb}$ space (Fig. 2a), most of the data from individual volcanoes define distinct linear Pb isotopic arrays, resolved outside the analytical error ($\sim 100 \text{ p.p.m.}$), indicating that there exists a significant heterogeneity in the components present within each of the two trends. The high $^{208}\text{Pb}/^{204}\text{Pb}$ ratios of Loa volcanoes in general, and Koolau and Lanai in particular, indicate the presence of an EM-1 component³⁰, the nature of which is discussed in the literature³¹. Here we note that these data rule out possible interpretations whereby Hawaiian melts represent simple two-component mixtures³². This is supported by both the HSDP-2 results, which, even when taken alone, indicate the presence of four distinct Pb components¹¹, and the distinct slopes of the individual Hawaiian Pb isotope arrays (Supplementary Table S2 and Fig. S1).

We performed a statistical analysis of the Pb isotope data (Supplementary Table S3) demonstrating that the populations of the Kea and Loa trends are significantly different from the pooled population at a significance level higher than 99.9% in $^{208}\text{Pb}/^{204}\text{Pb}$ – $^{206}\text{Pb}/^{204}\text{Pb}$ space. The regression analyses of the Pb isotope data for individual volcanoes are reported in Supplementary Table S2 and Fig. S1.

The Pb isotopic compositions of the two trends are clearly distinct, with overlap restricted to the anomalous samples from the submarine section of the HSDP-2 drill core (Kea-hi8, yellow diamonds in Fig. 1). However, in the absence of drill holes other than HSDP, sampling is necessarily limited. Nevertheless, the Loa–Kea differences exist regardless of the location of the volcanoes with respect to the plume centre. Our data for Loa trend volcanoes include early (Loihi), main- and late-shield (Mauna Loa) volcanism. Data for Kea volcanoes include main-shield (central location; old Mauna Kea and recent Kilauea) and late-shield (peripheral location; youngest Mauna Kea and Kohala) volcanism. We therefore argue that these data rule out previous interpretations in which the distinction between the Loa and Kea chains has been ascribed to the more central position of the Loa chain relative to the plume centre, and a more peripheral track of the Kea chain volcanoes.

In Fig. 2b, the data are plotted in $^{208}\text{Pb}^*/^{206}\text{Pb}^*$ – ϵ_{Nd} space, together with recently published Pb isotope data on Koolau landslide blocks³³, which are of similar precision to those using the Pb triple-spike technique, the $^{208}\text{Pb}^*/^{206}\text{Pb}^*$ ratio being used as a distinction criterion between the Loa trend and Kea trend. Despite the fact that the data form a broad negative correlation, the boundary between Loa (blue) and Kea (red) trends can be pinned down at a $^{208}\text{Pb}^*/^{206}\text{Pb}^*$ value of about 0.95, which is similar to that inferred from the slope of the Loa–Kea boundary in $^{208}\text{Pb}/^{204}\text{Pb}$ – $^{206}\text{Pb}/^{204}\text{Pb}$ space (Fig. 2a). Some overlap does occur across the $^{208}\text{Pb}^*/^{206}\text{Pb}^*$ boundary—in particular, the Kea-hi8

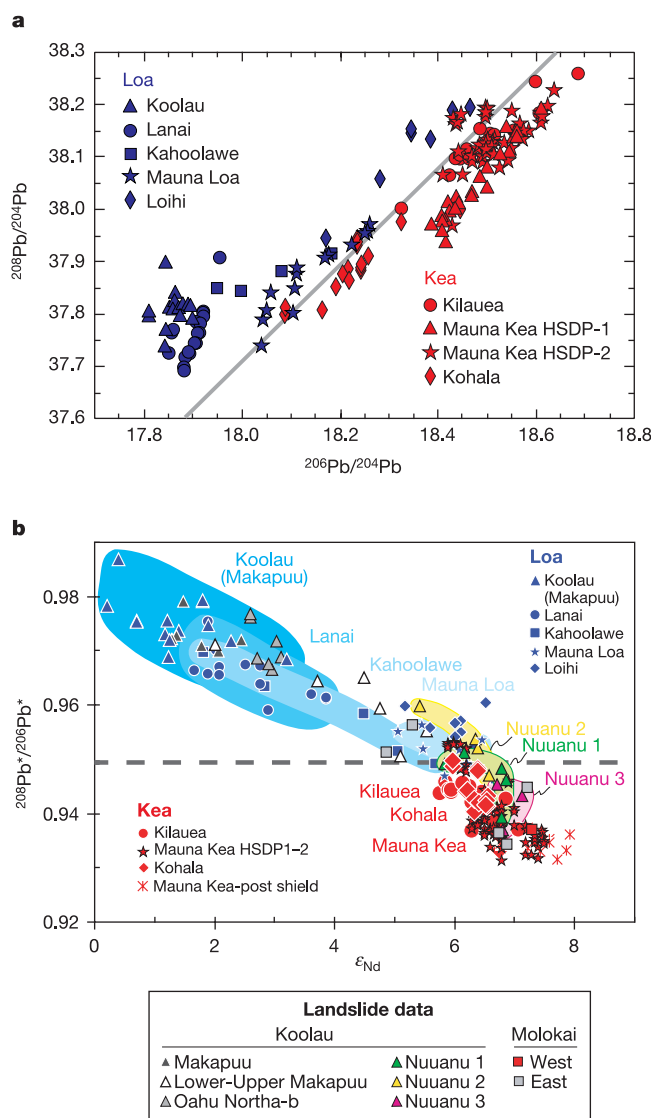


Figure 2 The Loa–Kea compositional division of the Hawaiian plume. **a**, Triple-spike Pb isotope data of Hawaiian shield stage lavas (except Loihi, which are early-shield stage) plotted in $^{208}\text{Pb}/^{204}\text{Pb}$ versus $^{206}\text{Pb}/^{204}\text{Pb}$ space. The grey line marks the boundary between Kea and Loa volcanoes, with higher $^{208}\text{Pb}/^{204}\text{Pb}$ ratios for Loa relative to Kea volcanoes at a given $^{206}\text{Pb}/^{204}\text{Pb}$ ratio. The only exceptions are a few lava flows from deeper levels in the Mauna Kea HSDP-2 core (red stars) which have Loihi-like features. Error bars (2s.d. = 100 p.p.m.) on Pb isotope ratios are similar to symbol size. **b**, ϵ_{Nd} versus $^{208}\text{Pb}^*/^{206}\text{Pb}^*$ ratios in Hawaiian volcanoes. Also plotted are the data for submarine landslide blocks from Koolau and Molokai³³. The grey dashed horizontal line marks the $^{208}\text{Pb}^*/^{206}\text{Pb}^*$ boundary value between Loa- (blue) and Kea- (red) trend volcanoes. Notable exceptions include Kea-hi8 lavas (red stars on Loa side) and Nuuuanu 3 (Koolau) landslide blocks, which partly fall on the Kea side.

samples from Mauna Kea plot on the Loa side and the Koolau landslide blocks partly fall on the Kea side of the boundary.

The greater Pb isotope compositional variability within the Loa trend (a factor of ~ 3) compared with that observed along the Kea trend (Fig. 2b) suggests larger source heterogeneities within the left (Loa) side compared with the right (Kea) side of the Hawaiian plume. The progressive increase in $^{208}\text{Pb}^*/^{206}\text{Pb}^*$ ratios and decrease in ϵ_{Nd} along the Loa chain with increasing volcano age further imply that sampling of the Loa heterogeneities is time-dependent. This is also reflected in the isotopic variations found in Koolau lavas (subaerial and submarine landslide blocks), which span the entire range found in Hawaiian volcanoes³³. There also appears to be a temporal evolution in Koolau Volcano from 'old', Kea-like to 'young', Loa-like isotope characteristics, although the

significance of this feature relies upon the validity of the reconstruction of the landslides, their origin and especially the relative ages between and within the Nuuanu landslide blocks (see ref. 33 and references therein). Recent Pb isotope data (our unpublished work) on Koolau Scientific Drilling Project (KSDP) lavas still show a consistent 'Loa' signature in the deeper stratigraphic levels of Koolau Volcano³⁴, suggesting that caution should be exercised in interpreting the landslide blocks data.

Timing and location of the Loa–Kea boundary

Using the $^{208}\text{Pb}^*/^{206}\text{Pb}^*$ ratio as a distinction criterion between the Loa and Kea trends and considering the geographic distribution of this parameter (Fig. 3a), we show that the extension of the Loa and Kea trends terminates at the Molokai fracture zone. In particular,

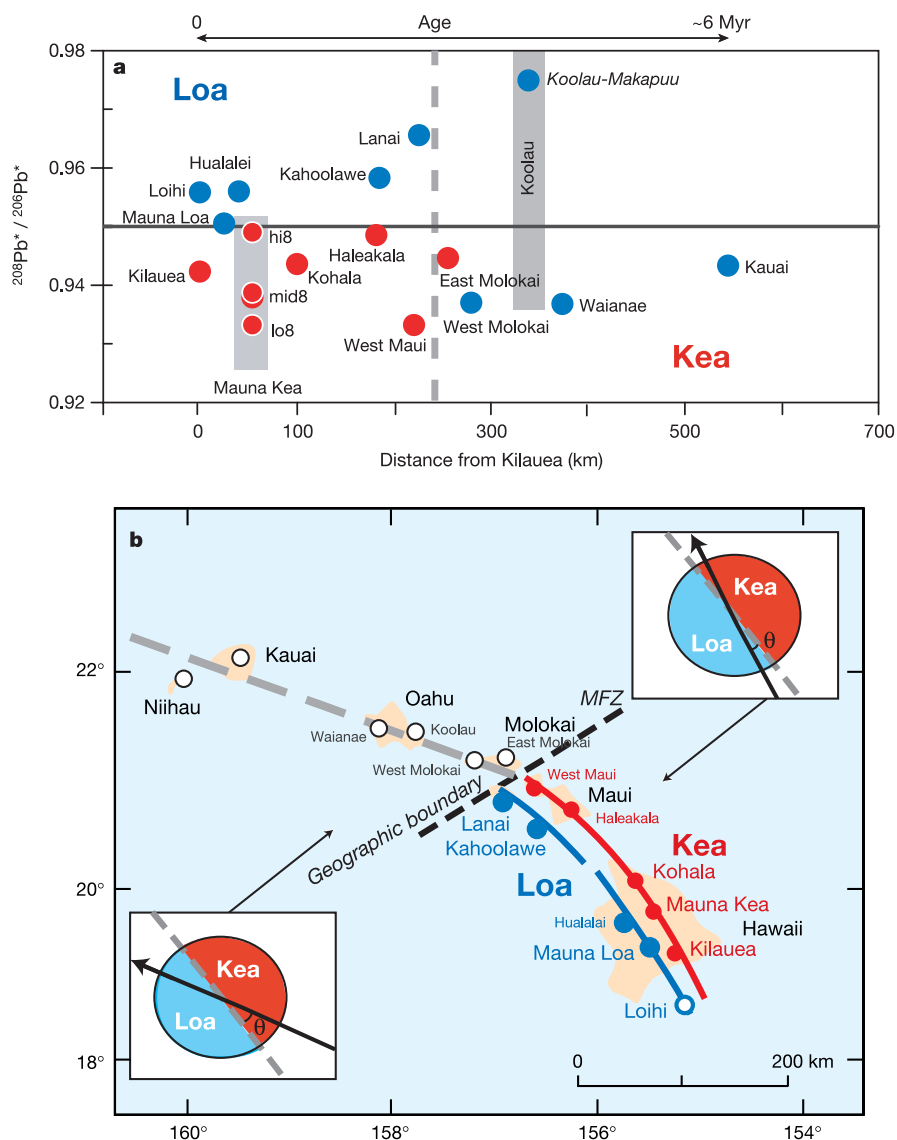


Figure 3 Geographic extension of the Loa–Kea compositional boundary. **a**, $^{208}\text{Pb}^*/^{206}\text{Pb}^*$ mean values versus distance from Kilauea for Hawaiian volcanoes. The upper x axis gives a corresponding timescale. Note that the Koolau data, shown as a grey shaded box³³, span the entire range of $^{208}\text{Pb}^*/^{206}\text{Pb}^*$ displayed by Hawaiian volcanoes. The blue circle labelled 'Koolau' corresponds to the mean $^{208}\text{Pb}^*/^{206}\text{Pb}^*$ of subaerial Makaapu stage lavas. The Loa–Kea distinction is disturbed beyond Lanai, as shown by the vertical dashed line. West Molokai, Waianae, Koolau (main-shield and landslide blocks) and Kauai, generally referred to as Loa chain volcanoes, have Kea-type $^{208}\text{Pb}^*/^{206}\text{Pb}^*$ ratios. Data are from this study and the GEOROC database. **b**, Map of the Hawaiian islands showing

that the distinct Loa–Kea chains, based on the spatial distribution of the $^{208}\text{Pb}^*/^{206}\text{Pb}^*$ ratio, terminate at the Molokai fracture zone, shown by the dashed black line. Beyond this geographic boundary, a single line passes through the volcanoes (black open circles) indicating that the Loa–Kea subdivision does not hold. The insets are schematic drawings illustrating the change in angle between the plate motion direction (solid black arrow) and the compositional Loa–Kea boundary (dashed grey line). The lower left inset corresponds to the case of Koolau volcano (2–3 Myr ago) and the upper right inset to the modern situation.

some of the main shield-stage lavas from Oahu (Koolau and Waianae) and Kauai volcanoes do not have the high $^{208}\text{Pb}^*/^{206}\text{Pb}^*$ ratios typical of Loa trend volcanoes (Fig. 3a), but rather have Kea-type features. We thus propose revising the geographic location of the Loa and Kea trend lines on the basis of the Loa–Kea compositional Pb isotopic boundary, which we infer does not extend beyond Lanai and Haleakala (Fig. 3b). Although this may appear to be at odds with the age-progression map³⁵, the coincidence of the compositional Pb isotopic transition with the bend of the Hawaiian islands chain near Molokai may be significant.

Geometric relocation of hotspots in the Pacific indicates that the Loa–Kea chains may be a recent feature that developed only about 2–3 Myr ago³⁶. This feature may have resulted from a change in the plate motion, as indicated by a bend of the Hawaiian chain near Molokai, and the effects of lithospheric flexure due to a volcanic load placed off-axis from the hotspot³⁷. The geographic location and the timing of the compositional Pb isotopic change agree well with these models, once the age uncertainties have been taken into account. Thus, the sampling of two compositionally different sides of the Hawaiian plume, with the left (Loa) side characterized by high $^{208}\text{Pb}^*/^{206}\text{Pb}^*$ and the right (Kea) side by low $^{208}\text{Pb}^*/^{206}\text{Pb}^*$, reflects the asymmetry of the Hawaiian plume. Such an asymmetry can be perpetuated over long distances and timescales by a dual line of volcanoes—with volcano spacing dependent upon lithospheric thickness³⁸—developing in response to changes in the direction of plate motion³⁷.

An explanation is required, however, for the isotopic transition of Koolau from a Kea-type signature (in the early shield-building stage) to a Loa-type signature (in the late shield-building stage)³³. Provided that the reconstruction and relative ages of the landslide blocks are valid³³, this transition can be explained if the azimuth of the compositional boundary between Loa and Kea-type plume chemistry is at a slight angle to the azimuth of the plate motion. Figure 3b illustrates how this may determine which side of the compositional boundary is sampled during a given evolutionary stage of a volcano. In the case of Koolau (left inset), the switch from Kea- to Loa-like chemistry implies that the angle between the vectors defined by the Loa–Kea compositional boundary and the plate motion was greater than the present-day angle (right inset). Such an explanation is reasonable given the uncertainties over the velocity of the Pacific plate relative to the plume and the direction of plate motion over the past 2 Myr (ref. 25). It is also quite possible that the change in the direction of plate motion 2–3 Myr ago may also have caused a change in the position of the Koolau volcano in such a way that its magma supply switched from one side of the plume to the other.

Bilateral asymmetry of the plume

The overall consistency of the Pb isotopic signature shown by the Loa and Kea chains, together with small-scale heterogeneities evident in the stratigraphic section of both Mauna Kea and Koolau, rules out models ascribing the compositional variations to purely concentric zoning of the plume^{6–9,39}. Instead, we infer that the plume possesses a gross lateral zonation (Supplementary Fig. S2). This model is consistent with the numerical simulations of plume evolution of Farnetani *et al.*¹³, showing that plumes draw compositional heterogeneities from a low-viscosity source layer into the plume stem by laminar flow, which should preserve any incidental, lateral, initial heterogeneities of the source layer. This source layer (perhaps located above the core–mantle boundary) is inferred to contain large-scale heterogeneities extending to perhaps hundreds of kilometres⁴⁰, which are drawn into, compressed and vertically stretched in the plume stem, and ultimately sampled by the two parallel trends on the Hawaiian islands. It is important, however, to emphasize that the lateral variations in the Hawaiian plume Pb isotope chemistry represent first-order observations and are thus independent of plume structure models. Rather, our data

demonstrate the existence of different scales of heterogeneities which, assuming a plate velocity of 10 cm yr^{-1} and upwelling velocities of 1 myr^{-1} , can be translated into length scales varying from tens of kilometres for the short-term (10 kyr) oscillations in HSDP-2 core to $\sim 50\text{ km}$ (Loa–Kea difference). Consequently, these results are inconsistent with a purely concentric zonation of the Hawaiian plume, as inferred from the He (and Nd) isotope distribution in the context of the dynamical plume structure model of refs 25 and 41. Recent three-dimensional numerical results show that radial variations in vertical velocity and induced shear stress across plume tails can generate long and distinct filaments leading to laterally heterogeneous rather than concentrically zoned plume tails⁴². In addition, fluid-dynamical tank experiments demonstrate that any left–right geochemical asymmetry in the plume source’s region will be preserved and reflected in the geochemistry of the erupted lavas⁴³. Altogether, these results provide strong support for our conclusion, which is essentially based on the Pb isotope anatomy of the Hawaiian plume.

Entrainment of mantle material overlying the plume-source layer²⁶ should lead to a concentric zonation, but this pattern is not reflected in our Pb isotope data. However, our sampling does not as yet include very-late-stage, post-erosional lavas that might be derived from the outermost parts of the plume and might therefore represent entrained ‘ambient’ mantle material. Therefore, we cannot firmly rule out the presence of entrained material in the outermost parts of the plume, but we can exclude concentric entrainment as the prime mechanism for explaining the occurrence of the Loa and Kea chains.

High-precision Pb isotope data on the scale of the Hawaiian islands chain demonstrate a large-scale left–right asymmetry in the Hawaiian plume. This asymmetry is reflected in the spatial distribution of the volcanoes along the Loa–Kea chain. Smaller-scale heterogeneities, revealed in the temporal records of Mauna Kea and Koolau lavas, also exist in the Hawaiian plume. These heterogeneities appear to be stretched out vertically within the plume and are sufficiently long-lived to be sampled by successive volcanoes. To progress beyond these general conclusions and develop a more detailed model of the plume anatomy, new high-resolution dynamic modelling will have to be carried out, incorporating the various scales of isotopic plume heterogeneity. □

Received 30 March 2004; accepted 16 January 2005; doi:10.1038/nature03402.

1. Dana, J. D. *Geology, United States Exploring Expedition, during the years 1838–1839, 1840, 1841, 1842* Vol 10 (C. Sherman, Philadelphia, 1849).
2. Jackson, E. D., Shaw, H. R. & Bargar, K. E. Calculated geochronology and stress field orientations along the Hawaiian chain. *Earth Planet. Sci. Lett.* **26**, 145–155 (1975).
3. Dalrymple, G. B., Silver, E. A. & Jackson, E. D. Origin of the Hawaiian islands. *Am. Sci.* **61**, 294–308 (1973).
4. Stille, P., Unruh, D. M. & Tatsumoto, M. Pb, Sr, Nd and Hf isotopic evidence of multiple sources for Oahu, Hawaii basalts. *Nature* **304**, 25–29 (1983).
5. Tatsumoto, M. Isotopic composition of lead in oceanic basalt and its implication to mantle evolution. *Earth Planet. Sci. Lett.* **38**, 63–87 (1978).
6. Hauri, E., Lassiter, J. C. & DePaolo, D. J. Osmium isotope systematics of drilled lavas from Mauna Loa, Hawaii. *J. Geophys. Res.* **101**, 11793–11806 (1996).
7. Kurz, M. D., Kenna, T. C., Lassiter, J. C. & DePaolo, D. J. Helium isotopic evolution of Mauna Kea Volcano: first results from the 1-km drill core. *J. Geophys. Res.* **101**, 11781–11791 (1996).
8. Lassiter, J. C., DePaolo, D. J. & Tatsumoto, M. Isotopic evolution of Mauna Kea volcano: Results from the initial phase of the Hawaiian Scientific Drilling Project. *J. Geophys. Res.* **101**, 11769–11780 (1996).
9. DePaolo, D. J. High-frequency isotopic variations in the Mauna Kea tholeiitic basalt sequence: melt zone dispersivity and chromatography. *J. Geophys. Res.* **101**, 11855–11864 (1996).
10. Sharp, W. A. & Renne, P. R. $^{40}\text{Ar}/^{39}\text{Ar}$ dating of core recovered by the Hawaii Scientific Drilling Project (phase 2), Hilo, Hawaii. *Geochem. Geophys. Geosyst.* (in the press).
11. Eisele, J., Abouchami, W., Galer, S. J. G. & Hofmann, A. W. The 320 kyr Pb isotope evolution of Mauna Kea lavas recorded in the HSDP-2 drill core. *Geochem. Geophys. Geosyst.* **4**, doi:10.1029/2002GC000339 (2003).
12. Blichert-Toft, J., Weis, D., Maerschalk, C., Agranier, A. & Albaredo, F. Hawaiian hot spot dynamics as inferred from the Hf and Pb isotope evolution of Mauna Kea volcano. *Geochem. Geophys. Geosyst.* **4**, 8704, doi:10.1029/2002GC000340 (2003).
13. Farnetani, C. G., Legras, B. & Tackley, P. J. Mixing and deformation in mantle plumes. *Earth Planet. Sci. Lett.* **196**, 1–15 (2002).
14. Galer, S. J. G. Optimal double and triple spiking for high precision lead isotopic measurement. *Chem. Geol.* **157**, 255–274 (1999).
15. Abouchami, W., Galer, S. J. G. & Hofmann, A. W. High precision lead isotope systematics of lavas from the Hawaiian Scientific Drilling Project. *Chem. Geol.* **169**, 187–209 (2000).

16. Stolper, E. M., Sherman, S., Garcia, M. O., Baker, M. & Seaman, C. Glasses in the submarine section of the HSDP-2 drill core, Hilo, Hawaii. *Geochem. Geophys. Geosyst.* **5**, doi:10.1029/2003GC000553 (2004).
17. Kurz, M. D., Curtice, J., Lott, D. E. & Solow, A. Rapid helium isotopic variability in Mauna Kea shield lavas from the Hawaiian Scientific Drilling Project. *Geochem. Geophys. Geosyst.* **5**, doi:10.1029/2002GC000439 (2004).
18. Bryce, J. G. & DePaolo, D. J. Sr and Nd variations in Mauna Kea lavas: preliminary results from analyses from the 1999 Hawaiian Scientific Drilling Project. *Eos Trans. AGU* **81**, abstr. V12C-01 (2000).
19. Galer, S. J. G. & O'Nions, R. K. Residence time of thorium, uranium and lead in the mantle with implications for mantle convection. *Nature* **316**, 778–782 (1985).
20. Garcia, M. O., Muenow, D. W. & Aggrey, K. E. Major element, volatile, and stable isotope geochemistry of Hawaiian submarine tholeiitic glasses. *J. Geophys. Res.* **94**, 10525–10538 (1989).
21. Yang, H. J., Frey, F. A., Garcia, M. O. & Clague, D. A. Submarine lavas from Mauna Kea Volcano, Hawaii: implications for Hawaiian shield stage processes. *J. Geophys. Res.* **99**, 15577–15594 (1994).
22. Easton, R. M. in *Volcanism in Hawaii* (eds Decker, R. W., Wright, T. L. & Stauffer, P. H.) 243–260 (US Geological Survey Prof. Pap. 1350, Boulder, Colorado, 1987).
23. Quane, S. L., Garcia, M. O., Guillou, H. & Hulsebosch, T. P. Magmatic history of the East Rift Zone of Kilauea Volcano, Hawaii based on drill core from SOH 1. *J. Volcanol. Geotherm. Res.* **102**, 319–338 (2000).
24. Lipman, P. W., Sisson, T. W., Ui, T., Naka, J. & Smith, J. R. in *Hawaiian Volcanoes: Deep Underwater Perspectives* (eds Takahashi, E., Lipman, P. W., Garcia, M. O., Naka, J. & Aramaki, S.) 161–191 (AGU, Washington, DC, 2002).
25. DePaolo, D. J., Bryce, J. G., Dodson, A., Shuster, D. L. & Mack Kennedy, B. Isotopic evolution of Mauna Loa and the chemical structure of the Hawaiian plume. *Geochem. Geophys. Geosyst.* **2**, doi:10.1029/2000GC000139 (2001).
26. Hauri, E. H., Whitehead, J. A. & Hart, S. R. Fluid dynamic and geochemical aspects of entrainment in mantle plumes. *J. Geophys. Res.* **99**, 24275–24300 (1994).
27. Loper, D. E. & Stacey, F. D. The dynamical and thermal structure of deep mantle plumes. *Phys. Earth Planet. Inter.* **33**, 304–317 (1983).
28. Olson, P., Schubert, G. & Anderson, C. Structure of axisymmetric mantle plumes. *J. Geophys. Res.* **98**, 6829–6844 (1993).
29. Schubert, G., Turcotte, D. L. & Olson, P. *Mantle Convection in the Earth and Planets* (Cambridge Univ. Press, New York, 2001).
30. Zindler, A. & Hart, S. R. Chemical geodynamics. *Annu. Rev. Earth Planet. Sci.* **14**, 493–571 (1986).
31. Eisele, J. *et al.* The role of sediment recycling in EM-1 inferred from Os, Pb, Hf, Nd, Sr isotope and trace element systematics of the Pitcairn hotspot. *Earth Planet. Sci. Lett.* **196**, 197–212 (2002).
32. Blichert-Toft, J., Albarède, F. & Frey, F. Hf isotopic evidence for pelagic sediments in the source of Hawaiian basalts. *Science* **285**, 879–882 (1999).
33. Tanaka, R. & Nakamura, E. in *Hawaiian Volcanoes: Deep Underwater Perspectives* (eds Takahashi, E., Lipman, P. W., Garcia, M. O., Naka, J. & Aramaki, S.) 311–332 (AGU, Washington, DC, 2002).
34. Fekiacova, Z. & Abouchami, W. Pb isotopic evolution of Koolau volcano (Oahu, Hawaii). *Eos Trans. AGU* **84**, V32A-0991 (2003).
35. Clague, D. A. & Dalrymple, G. B. in *The Eastern Pacific Ocean and Hawaii* (eds Winterer, E. L., Hussong, D. M. & Decker, R. W.) 188–217 (Geological Society of America, Boulder, Colorado, 1989).
36. Wessel, P. & Kroenke, L. A geometric technique for relocating hotspots and refining absolute plate motions. *Nature* **387**, 365–369 (1997).
37. Hieronymus, C. F. & Bercovici, D. Discrete alternating hotspot islands formed by interaction of magma transport and lithospheric flexure. *Nature* **397**, 604–607 (1999).
38. ten Brink, U. Volcano spacing and plate rigidity. *Geology* **19**, 397–400 (1991).
39. Hauri, E. H. Major-element variability in the Hawaiian mantle plume. *Nature* **382**, 415–419 (1996).
40. Kellog, L. H., Hager, B. H. & van der Hilst, R. V. Compositional stratification in the deep mantle. *Science* **283**, 1881–1884 (1999).
41. Ribe, N. M. & Christensen, U. R. The dynamical origin of Hawaiian volcanism. *Earth Planet. Sci. Lett.* **171**, 517–531 (1999).
42. Farnetani, C. & Samuel, H. Dynamics of thermochemical plumes. *Eos Trans. AGU* **85**, V44B-03 (2004).
43. Kerr, R. C. & Meriaux, C. Structure and dynamics of sheared mantle plumes. *Geochem. Geophys. Geosyst.* **5**, doi:10.1029/2004GC000749 (2004).

Supplementary Information accompanies the paper on www.nature.com/nature.

Acknowledgements We thank M. Garcia, H. West, and the HSDP team for providing samples, and J. Bryce for providing updated Nd isotope data for HSDP-2. Comments from M. Garcia and E. Takahashi on Nuanu landslides stratigraphy, and reviews from E. Hauri and B. Hanan were appreciated.

Competing interests statement The authors declare that they have no competing financial interests.

Correspondence and requests for materials should be addressed to W.A. (wafa@mpch-mainz.mpg.de).

A common sex-dependent mutation in a *RET* enhancer underlies Hirschsprung disease risk

Eileen Sproat Emison^{1*}, Andrew S. McCallion^{1*}, Carl S. Kashuk¹, Richard T. Bush¹, Elizabeth Grice¹, Shin Lin¹, Matthew E. Portnoy², David J. Cutler¹, Eric D. Green^{2,3} & Aravinda Chakravarti¹

¹McKusick – Nathans Institute of Genetic Medicine, Johns Hopkins University School of Medicine, Baltimore, Maryland 21205, USA

²Genome Technology Branch and ³NIH Intramural Sequencing Center, National Human Genome Research Institute, National Institutes of Health, Bethesda, Maryland 20892, USA

* These authors contributed equally to this work

The identification of common variants that contribute to the genesis of human inherited disorders remains a significant challenge. Hirschsprung disease (HSCR) is a multifactorial, non-mendelian disorder in which rare high-penetrance coding sequence mutations in the receptor tyrosine kinase *RET* contribute to risk in combination with mutations at other genes. We have used family-based association studies to identify a disease interval, and integrated this with comparative and functional genomic analysis to prioritize conserved and functional elements within which mutations can be sought. We now show that a common non-coding *RET* variant within a conserved enhancer-like sequence in intron 1 is significantly associated with HSCR susceptibility and makes a 20-fold greater contribution to risk than rare alleles do. This mutation reduces *in vitro* enhancer activity markedly, has low penetrance, has different genetic effects in males and females, and explains several features of the complex inheritance pattern of HSCR. Thus, common low-penetrance variants, identified by association studies, can underlie both common and rare diseases.

HSCR, or congenital aganglionosis with megacolon, occurs in one in every 5,000 live births. Heritability of HSCR is nearly 100% with clear multigenic inheritance. Although *RET* is the main HSCR gene implicated^{1,2}, mutations also occur in seven other genes involved in enteric development, specifically *ECE1*, *EDN3*, *EDNRB*, *GDNF*, *NRTN*, *SOX10* and *ZFHX1B*³. However, less than 30% of patients have mutations in these eight genes; additional HSCR-causing mutations in *RET* and/or other genes must therefore exist.

We recently identified a high-frequency HSCR-associated *RET* haplotype in an inbred Mennonite population⁴. However, this haplotype lacks a frank coding-sequence mutation. Association between HSCR and specific *RET* marker alleles has also been shown in other populations^{5–7}. Importantly, even in families with established linkage to the *RET* locus on chromosome 10q11.21, *RET* coding mutations are frequently absent^{1,2}. Additionally, we demonstrated that the *RET* Mennonite haplotype is also present in the general population and is overtransmitted to affected offspring⁸, indicating that the same mutation might underlie HSCR in Mennonites and in the general outbred population. We advance two competing, but not mutually exclusive, hypotheses to explain these findings: first, one or more non-coding mutations at *RET*, and second, a mutation(s) in a second gene that is tightly linked to *RET*. The shared conjecture of these hypotheses is that a common HSCR-causing mutation lies in a functional sequence within or flanking the *RET* locus. Resolving these hypotheses can be accomplished by a congruence of association studies and a comprehensive catalogue of all functional elements within the greater *RET* locus.

Family-based association studies

Genome sequence data (<http://www.genome.ucsc.edu/cgi-bin/hgGateway>; build 35) identifies two additional genes in the 350-kilobase (kb) region surrounding *RET*. *GALNACT-2*, a chondroitin *N*-acetylgalactosaminyltransferase^{9,10}, contains eight exons spanning 46.8 kb and begins 9 kb from the last *RET* exon.

Thirteen exons encode *RASGEF1A*, a predicted guanyl-nucleotide exchange factor that spans 72 kb and begins 65 kb 3' of *RET*. To refine the association within this locus genetically, we initially genotyped 28 single-nucleotide polymorphisms (SNPs) spanning 175 kb in 126 HSCR-affected individuals and their parents, ascertained from the general outbred population (Table 1). The genomic interval encompasses *RET*, *GALNACT-2* and *RASGEF1A*.

Transmission disequilibrium tests (TDTs)¹¹ on each SNP showed statistically significant disease associations spanning a region immediately 5' of *RET* through to *RASGEF1A* (Fig. 1a, Table 1). Specifically, 13 of 17 *RET* SNPs, 3 of 7 *GALNACT-2* SNPs and 2 of 4 *RASGEF1A* SNPs tested are significantly associated with HSCR (Table 1), reflecting the high background linkage disequilibrium in this region (data not shown). However, the greatest statistical significance, and, more importantly, the largest transmission distortions ($\tau \geq 0.7$), occurred among eight SNPs in a 27.6-kb segment from 4.2 kb 5' of *RET* through to *RET* exon 2 (Fig. 1a). Within this region the highest association was within *RET* intron 1.

We performed and analysed three resequencing experiments to identify additional variants, with particular emphasis being given to multi-species conserved sequences (MCSs; see below) within the 27.6-kb region of highest association. Specifically, we identified the SNP RET+3 (marked with an asterisk in Fig. 1a) within MCS+9.7 by resequencing HSCR patients from families with demonstrated *RET*-linkage but no identified coding sequence mutations. TDTs of RET+3 in all 126 trios showed the largest transmission distortion ($\tau = 0.8$) and the highest statistical significance ($P = 10^{-11}$). When association tests are factored by offspring gender, a known risk factor in HSCR, RET+3 and the adjacent marker 1.1SfCI (3.3 kb away) are the only two SNPs demonstrating association in females. Two additional variants (rs2506005 and rs2506004) lie within MCS+9.7 and are located 76 nucleotides (nt) 5' and 217 nt 3' of RET+3, respectively; both are in complete linkage disequilibrium with RET+3 and each other. The HSCR-associated allele at each of

these additional SNPs is the ancestral allele. The RET+3:C allele is highly conserved in all nine mammalian species examined (Supplementary Fig. 1), and it is the derived polymorphic allele (RET+3:T) that is overtransmitted. We postulate that RET+3 is the most likely site of the disease variation.

We queried whether susceptibility to HSCR within this locus can be explained by *RET* alone or whether additional common variants might be present at *GALNACT-2* or *RASGEF1A*. We used the exhaustive allelic TDT (EATDT), a method for iteratively and successively testing all possible haplotypes of all possible sizes for association with HSCR^{12,13}. Seventeen haplotypes are significantly associated with HSCR, but they have two critical properties (Fig. 1b): first, no associated haplotype is limited to markers across *GALNACT-2* or *RASGEF1A*, and second, all haplotypes involve *RET* SNPs alone, particularly those in intron 1. These results strongly indicate a role for a single, common variant within *RET*. Because all but one haplotype involves RET+3, we conclude that the HSCR association can be attributed to one of the following: RET+3 is in tight linkage disequilibrium with a yet unknown disease-susceptibility variant, RET+3 is the disease-causing mutation alone, or RET+3 is a disease-causing variant that acts synergistically with additional disease variants on the associated haplotype.

Comparative genomics to define functional elements

The finding of association across an intron indicated a need to identify functional elements within the *RET* locus. Systematic comparisons of orthologous sequences can uncover coding and non-coding functional elements on the assumption that such regions evolve more slowly than non-functional (neutral) sequences^{14–17}. We obtained and compared the genomic sequence of a ~350-kb segment encompassing human *RET* with the ortho-

logous intervals in 12 non-human vertebrates. MCSs were identified as the intersection of elements that satisfied the criteria of Bray¹⁸ and Margulies¹⁹. Synteny is preserved across this interval in all vertebrates examined, although the fraction of sequence that can be aligned with the human sequence decreases with increasing evolutionary distance (Fig. 2a).

A total of 84 MCSs were identified (Supplementary Table 1), with 44% (37 of 84) of the identified MCSs corresponding to exons of *RET*, *GALNACT-2* and *RASGEF1A*. The remaining 47 MCSs are likely to be non-coding because no matching complementary DNA sequence or open reading frame greater than 20 amino acids in length was found. We identified five such elements within the most highly associated 27.6 kb around *RET* intron 1 (MCS–5.2, MCS–1.3, MCS+2.8, MCS+5.1 and MCS+9.7, identified by their distance in kilobases from the *RET* start site as in Fig. 3a).

Although *GALNACT-2* and *RASGEF1A* are unlikely to harbour common HSCR variants, they might carry rare mutations and be important in HSCR, just as some of the 126 patients we studied also have rare *RET* mutations. To test their involvement in enteric development and HSCR, we characterized their temporal and spatial expression in humans and mice. Transcription of *RASGEF1A* is limited to brain and several tissues (bone marrow, testis, colon and placenta) with high replicative capacity (Fig. 2b–d). *RET* and *GALNACT-2* share overlapping, nearly ubiquitous postnatal expression patterns. Importantly, *GALNACT-2* and *RASGEF1A* are both highly expressed at 13.5 days *post coitum*, coincident with peak *RET* expression and colonization of the gut by neural-crest-derived neuronal precursors (Fig. 2c), a feature disrupted in HSCR. Consequently, *GALNACT-2* and *RASGEF1A* expression patterns are consistent with a potential role in enteric neural crest migration. However, our analysis of morpholino-based gene

Table 1 Analysis of disease associations

Gene	Marker	dbSNP ID	A1	A2	All affected individuals			Male offspring			Female offspring		
					T	U	τ	T	U	τ	T	U	τ
5' <i>RET</i>	RET–6	rs3097566	G	T	45	43	0.51	27	32	0.46	18	11	0.62
	RET–5	rs2742250	G	C	56	44	0.56	41	27	0.60	15	17	0.47
	RET–4	rs3026707	A	G	45	33	0.58	32	19	0.63	13	14	0.48
	RET–3	rs3026720	T	C	38	29	0.57	27	17	0.61	11	12	0.48
	RET–2†	rs741763	G	C	69	26	0.73**	47	14	0.77**	22	12	0.65
	RET–1†	rs2505997	C	T	57	19	0.75**	43	10	0.81**	14	9	0.61
<i>RET</i> int1	RET+1†	rs2435365	T	C	76	29	0.72**	53	17	0.76**	23	12	0.66
	RET+2†	rs2435364	A	G	73	27	0.73**	50	15	0.77**	23	12	0.66
	1.1Sfcl†	rs2435362	A	C	100	27	0.79***	72	14	0.84***	28	13	0.68*
	RET+3††	rs2435357	T	C	101	25	0.80***	73	12	0.86***	28	13	0.68*
	RET+4†	rs752975	A	G	74	29	0.72**	51	17	0.75**	23	12	0.66
	INT1.4b†	rs2505535	G	A	92	28	0.77***	68	14	0.83***	24	14	0.63
<i>RET</i> protein-coding region	X2Eagl†	rs1800858	A	G	96	28	0.77***	72	14	0.84***	24	14	0.63
	INT8	rs3026750	G	A	60	40	0.60*	42	22	0.66*	18	18	0.50
	X13TaqI	rs1800861	G	T	59	38	0.61*	41	21	0.66*	18	17	0.51
	I18BbvI	rs2742237	C	G	59	28	0.68*	41	14	0.75**	18	14	0.56
	I18StyI	rs2742239	A	G	52	27	0.66*	34	12	0.74**	18	15	0.55
	I19BsgI	rs2075912	T	C	55	25	0.69*	37	12	0.76**	18	13	0.58
<i>GALNACT-2</i>	GN–1	rs3026787	G	A	17	14	0.55	15	10	0.60	2	4	0.33
	GN+1	rs4948705	C	T	59	29	0.67*	40	13	0.75**	19	16	0.54
	GN+2	rs1864393	A	G	35	15	0.70*	27	9	0.75**	8	6	0.57
	GN+3	rs2435337	G	C	57	29	0.66*	39	14	0.74**	18	15	0.55
	GN+4	rs2505556	C	T	63	59	0.52	42	39	0.52	21	20	0.51
	GN+5	rs2435384	G	T	57	39	0.59	39	21	0.65*	18	18	0.50
<i>RASGEF1A</i>	GN+6	rs2435381	T	C	55	41	0.57	37	28	0.57	18	13	0.58
	RAS+2	rs1254958	T	C	56	27	0.67*	38	13	0.75**	18	14	0.56
	RAS+1	rs1254965	T	C	56	27	0.67*	38	12	0.76**	18	15	0.55
	RAS–1	rs1272142	G	T	55	41	0.57	38	22	0.63*	17	19	0.47
	RAS–2	rs1955356	A	T	51	39	0.57	33	24	0.58	18	15	0.55

A1 and A2 are associated and unassociated alleles, respectively; T and U are transmitted and untransmitted allele counts, respectively; τ is transmission frequency of A1.

†SNPs highlighted by the black box on Fig. 1a and displayed in Fig. 3a.

‡rs2506005 and rs2506004 lie 76 nt upstream and 217 nt downstream of RET+3, respectively. These three SNPs were found to be in complete linkage disequilibrium; the TDT results are therefore identical with those for RET+3.

Significant *P* values are represented as follows: asterisk, $0.001 < P < 0.05$; two asterisks, $10^{-6} < P < 10^{-3}$; three asterisks, $10^{-11} < P < 10^{-9}$.

knockdowns in zebrafish has uncovered only mid-gastrulation defects in convergence and extension for *Galnact-2* and neuronal cell death in the central nervous system by 24 h after fertilization for *Rasgef1a* (data not shown). In contrast, a similar disruption of *RET* results in incomplete colonization of the digestive tube by enteric neurons^{20,21}. These functional analyses cannot exclude either *GALNACT-2* or *RASGEF1A* as HSCR candidate genes, because the observed embryonic lethality occurred before the onset of migration of neural crest cells into the digestive tube. However, genetic association tests have excluded the occurrence of a common mutation at *GALNACT-2* or *RASGEF1A* as a contributory factor in HSCR.

MCS+9.7 functions as an enhancer *in vitro*

Although MCS+9.7 is probably a functional element, the specific function of this sequence and the mechanism by which it exhibits a deleterious effect is not known. MCS+9.7 demonstrates a minimum identity of 72.5% with all mammalian species examined. No predicted structural or regulatory RNAs were identified in MCS+9.7 with the QRNA algorithm²². The MCS+9.7 sequence includes a gamut of predicted transcription-factor-binding sites (Supplementary Table 2), including two retinoic acid response elements (RARE) within 4 nt on each side of the RET+3 site. However, no predicted binding sites are disrupted directly by the

mutant RET+3:T allele or the alleles at the rs2506004 and rs2506005 sites. Retinoic acid has already been documented as a negative and a positive regulator of *RET* expression in cardiac and renal development, respectively^{23,24}. Furthermore, exogenous retinoic acid delays the colonization of the hindgut by *RET*-positive enteric neuroblasts and results in ectopic *RET* expression during embryogenesis²⁵. Although the mutation or mutations do not introduce or destroy a predicted RARE, it might introduce a new site that permits competition with, or reduces access to, the neighbouring predicted RAREs. Thus, the ultimate proof of disease causation will require the synthesis of the trait, from one or all three of the MCS+9.7 variants, in an appropriate model organism.

On the basis of its location, we predicted that the MCS+9.7 element functions as a transcriptional enhancer or suppressor. Using transient transfection assays, we tested the function of two *RET* intron 1 constructs in the mouse neuroblastoma cell line Neuro-2a. Amplicons containing MCS+9.7 and MCS+5.1/+9.7 show enhancer activity in this cell line (Fig. 3b), although this activity in HeLa cells is negligible (data not shown), indicating that the activity of MCS+9.7 might be dependent on cell type. Importantly, amplicons harbouring the mutant allele have significantly lower enhancer activity (sixfold to eightfold decrease) than those containing the wild-type allele (*t*-test, $P \leq 0.001$). These data indicate that the mutation might lie within, and might compromise

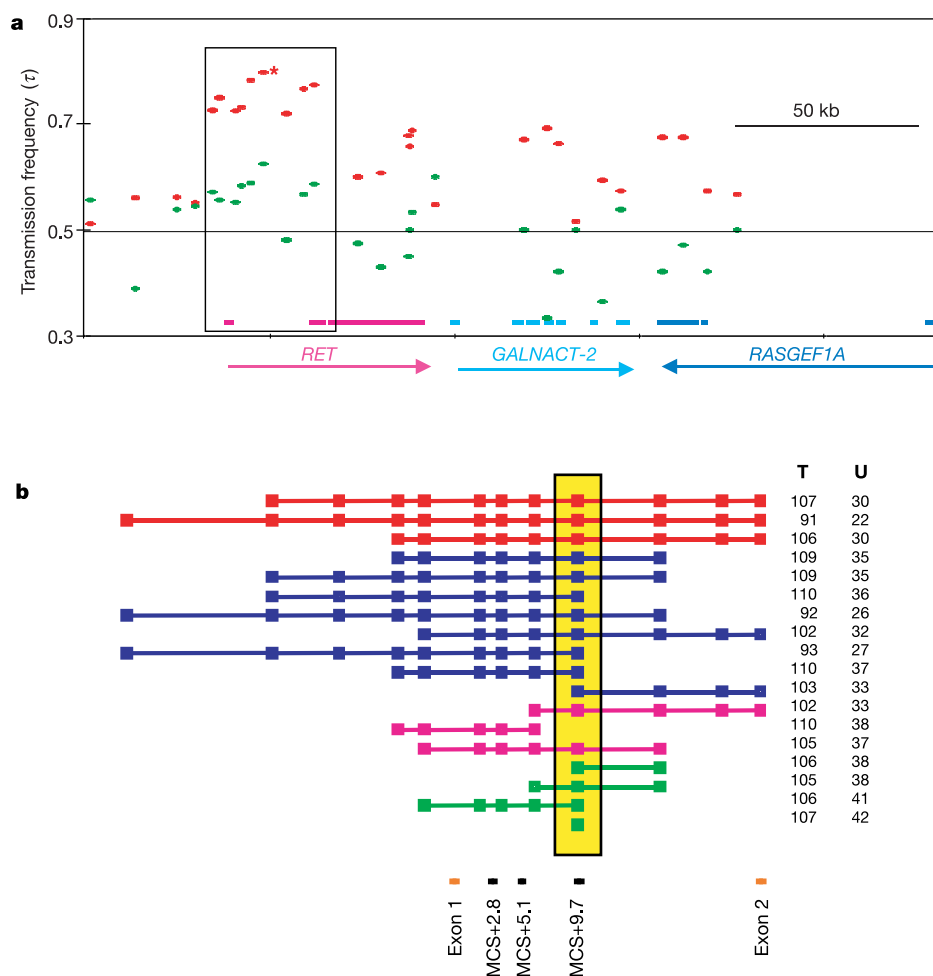


Figure 1 Transmission disequilibrium tests. **a**, TDT tests of individual SNPs. The region of 10q11.21 including *RET*, *GALNACT-2* and *RASGEF1A*. The horizontal line at 50% transmission indicates expectation under the null hypothesis. The asterisk identifies RET+3. Exons are marked as coloured boxes. The black rectangle represents the 27-kb area in Fig. 3a. Red symbols, affected offspring; green symbols, unaffected offspring.

b, EATDTs. The 5'-most SNP shown is RET-5, the 3'-most SNP is X2Eagl. Counts of transmitted (T) and untransmitted (U) chromosomes are given in columns at the right. All haplotypes with permutation-based *P* values less than or equal to the single most significantly associated SNP (RET+3) are shown. Red, $P < 10^{-8}$; blue, $P = 10^{-8}$; purple, $P = 10^{-7}$; green, $P = 10^{-6}$.

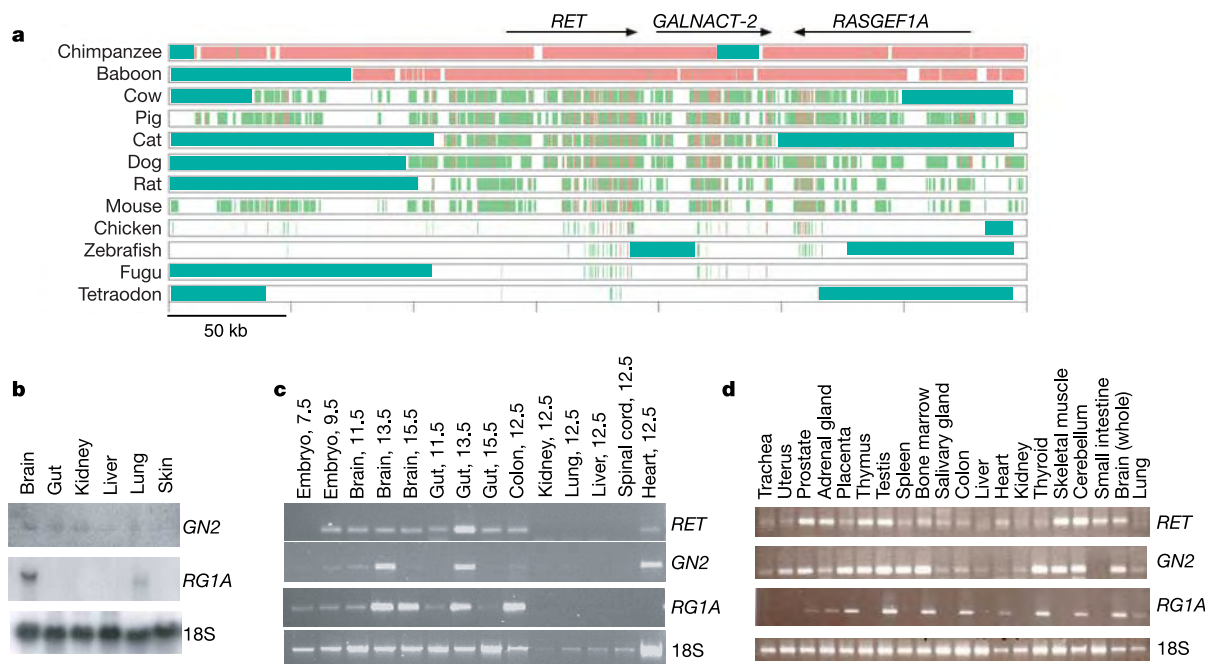


Figure 2 Identification and characterization of conserved sequence elements within 350 kb encompassing *RET*. **a**, Multi-PIP (per cent identity plot) alignment of genomic sequence from 12 vertebrates compared with the human sequence. Red, more than 75% sequence identity over 100 nt; green, more than 50% sequence identity over 100 nt; blue, gaps in contig of 500 nt or more; white, unalignable. **b**, Northern blots showing expression of *GALNACT-2* (GN2) and *RASGEF1A* (RG1A) in adult mouse tissues. **c**, **d**, Expression of *RET*, *GALNACT-2* (GN2) and *RASGEF1A* (RG1A) by RT-PCR in embryonic mouse (**c**) and adult human tissues (**d**). Numbers in **c** are days *post coitum*.

the activity of, an enhancer-like sequence in *RET* intron 1. *RET* coding sequence mutations in HSCR are always loss-of-function alleles. Thus, our finding that the RET+3 mutation decreases transcription is consistent with HSCR biology. We can localize the enhancer function, and the genetic change that diminishes that

function, to the 900-nt fragment tested in the MCS+9.7 construct. Within this region there exist three segregating sites (rs2506005, RET+3 and rs2506004) in complete linkage disequilibrium. In principle, any one of these three sites, or their combination, can be the disease susceptibility factor.

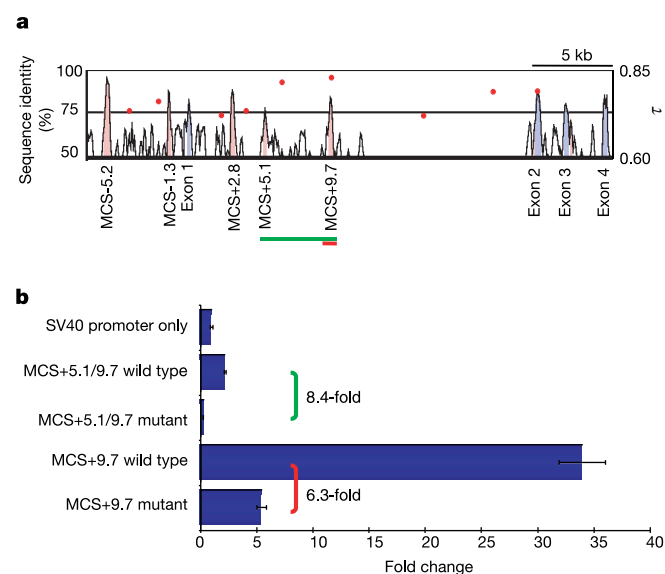


Figure 3 Identification and functional characterization of *RET* MCS+9.7. **a**, VISTA plot displaying percentage identity between mouse and human in the 5' region of *RET*. Estimated transmission frequencies to affected offspring are shown by red circles. **b**, Reporter gene expression in Neuro-2a cells with the use of amplicons MCS+9.7 and MCS+5.1/9.7 (mutant and wild type correspond to nucleotides T and C, respectively). The smaller of the tested constructs (MCS+9.7 only) is bracketed in red; the MCS+5.1/9.7 amplicon encompassing both MCS+9.7 and the adjacent MCS+5.1 is bracketed in green. All assays were conducted in triplicate and were repeated three times (nine data points total). Error bars represent standard error.

Worldwide distribution of MCS+9.7 variants

The global distribution of the RET+3:T allele was determined by genotyping individuals from 51 unselected populations. The mutant T allele is virtually absent within Africa (a frequency of less than 0.01), has intermediate frequency in Europe (0.25) but reaches high frequency (0.45) in Asia (Fig. 4). Additionally, we generated haplotypes for seven SNPs from 180 individuals, 60 each from Africa, Europe and Asia, derived from the above worldwide set, and compared them against haplotypes from HSCR patients (Supplementary Table 3). Haplotypes bearing the RET+3:T allele probably have a single origin, some time after modern humans emerged from Africa. The high frequency of the RET+3:T allele, and the susceptibility haplotype, in east Asia correlates with an increased incidence of short-segment HSCR among Asian newborns (3.1 versus 1.5 out of every 10,000 births in Asian American versus European American births in California between 1983 and 1997; C. Torfs, personal communication). This same haplotype has a 66% frequency in Chinese sporadic HSCR patients⁶; consequently, a doubling of the mutant allele frequency translates into an approximate doubling of disease incidence. We suspect that RET+3:T is a marker for short-segment HSCR because the low frequency of the RET+3:T allele in Africa correlates with a lower frequency of short-segment HSCR among African Americans³.

These data strongly suggest that, of the three SNPs within MCS+9.7, only the RET+3 variant is the susceptibility mutation. The associated alleles at rs2506005, RET+3 and rs2506004 are the ancestral, derived and ancestral alleles, respectively. Given our knowledge of human evolution and that the susceptibility haplotype has 1% frequency in Africa, the ancestral haplotype (with ancestral alleles at each SNP) was virtually extinct within Africa

until it increased in frequency with the occurrence of the RET+3:T mutation.

This finding of a common allele that rapidly increased in frequency but is associated with a disease predisposition can be explained in one of three ways: first, recurrent mutations from the wild type to the same deleterious mutant; second, chance increase by genetic drift; and third, a selective advantage of the mutation in heterozygotes. Our finding of a common haplotype indicates that the first explanation may be unlikely. To distinguish between the two remaining alternatives, we performed two analyses: we estimated an F_{ST} value of 0.027, and we compared our worldwide mutant allele distribution (summarized as an allele frequency of less than 5% in Africa, more than 25% in Europe, and more than 40% in China and Japan) to that of 8,247 SNPs from the ENCODE loci²⁶. Only 38 sites (0.46%) show the observed or a more extreme pattern, strongly indicating a selective advantage to the mutation.

We do not know the nature of the selective agent or whether such selection was only in the past or exists today. Common mutations and polymorphisms have survived for long periods of human evolution and have been exposed to numerous biological and demographic factors that have affected their survival and current frequency. If polymorphisms make substantial contributions to common disorders, a significant fraction of them must have been exposed to selection. It is not therefore surprising that most common disease associations involve alleles that provided (α -globin, β -globin²⁷, glucose-6-phosphate dehydrogenase G6PD^{28–30}, HLA³¹, Fy³² and other variants in malaria), or are suspected of providing (CCRΔ32 in HIV infection^{33–35}), a survival advantage to humans. Thus, many common variants in currently common disorders perhaps stem from alleles that were, or are, protective for another phenotype, providing mechanistic support to the common variant, common disease model of genetic disease^{36,37}.

Before the advent of corrective surgical methodologies in the 1950s, HSCR was a uniformly fatal disorder, necessitating positively acting selective forces to maintain this deleterious allele at high frequency. Our demonstration that the RET+3:T allele is a derived allele that is virtually absent in Africa but rose to a frequency of 0.25 in Europe and 0.45 in Asia in 100,000 years or less is indicative of such a selective force. RET is a tyrosine kinase receptor on the

surface of neuroblasts and many other cell types, and it is not inconceivable that it might be a target of pathogen entry, similar to the chemokine receptors involved in HIV and malaria.

Genetic properties of the RET+3 susceptibility allele

A pervasive feature of HSCR is the marked gender difference in expression and incidence, with males being four times more likely to be affected than females. These sex differences could arise from mutations on the X chromosome, but genome-wide mapping studies^{1,2} have consistently failed to identify an X-linked gene. Consequently, we tested whether the RET+3 variant at MCS+9.7 shows sex-specific effects. As shown in Table 1, transmission frequency of the associated allele in the RET region is always smaller to affected daughters than to affected sons, with rare exceptions at non-significant SNPs. Indeed, given the lower female penetrance, there were fewer affected daughters than sons in our sample, and among them only the mutant SNP (boys, $\tau = 0.86$, $P = 3.7 \times 10^{-11}$; girls, $\tau = 0.68$, $P = 0.02$) and the SNP at 1.1Sfcl, 3.3 kb away, are statistically significantly different from 0.50. Nevertheless, a trend test for a difference in transmission frequency in male and female offspring is highly significant and estimates the male-to-female transmission ratio to be approximately 2 ($P = 0.0007$). Thus, the genetic effect at MCS+9.7 is significantly greater in sons than in daughters.

Two other features of the RET+3 mutation display sex differences consistent with the greater incidence in males than females. First, as shown in Table 2, the transmission frequency to affected sons and daughters leads to a 5.7-fold and 2.1-fold increase in susceptibility in males and in females, respectively, assuming a multiplicative model for penetrance. Second, genotype frequencies of affected individuals can be used to estimate the penetrance, which varies between 6.2×10^{-5} and 1.8×10^{-3} (Table 2) and is considerably smaller than that for long-segment HSCR. Our finding of gender differences in penetrance is consistent with the greater incidence of HSCR in males. For all traits with gender-specific differences in incidence, affected individuals from the less frequently affected sex (females for HSCR) have a higher mean liability. Therefore, when we consider all susceptibility loci, we expect females with HSCR to carry more susceptibility alleles than their male counterparts³⁸. It

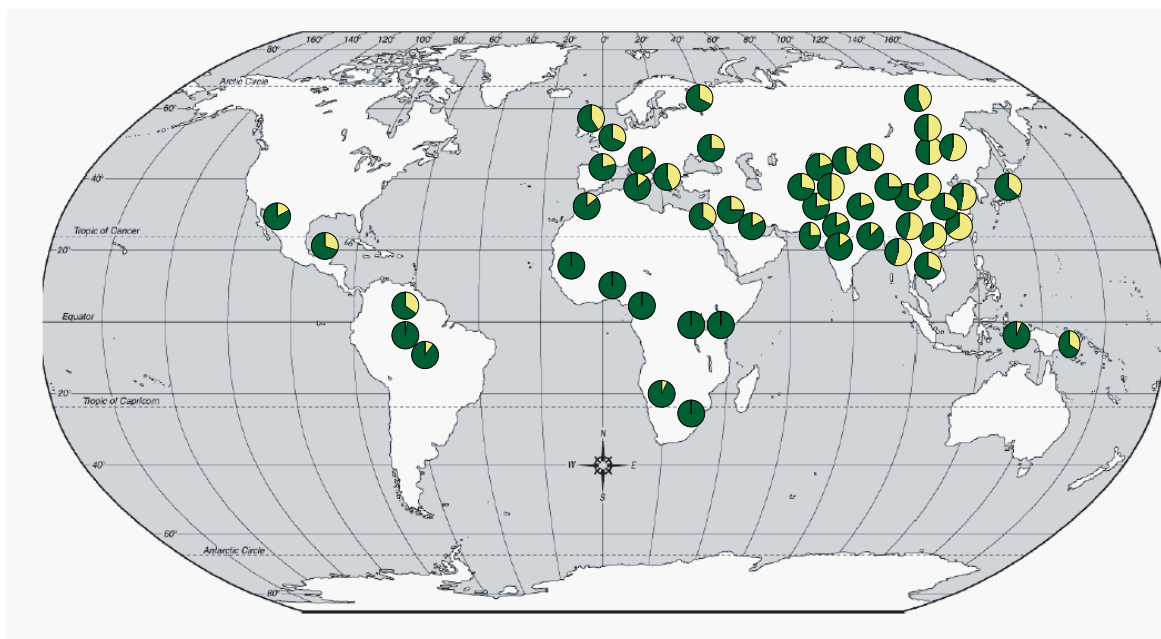


Figure 4 Worldwide allele frequencies of RET+3. Frequencies of the putative wild-type (green, C) and mutant (yellow, T) alleles are given for 51 populations comprising 1,064 individuals from the CEPH Human Genome Diversity Panel.

follows that the penetrance of any specific mutation must be lower for the less-affected sex, as observed here.

To assess the genetic impact of this common mutation we estimated the proportion of the total variance in susceptibility that the RET+3 mutation explains. Only 2.63% and 1.14% of the variation are explained by the action of this mutation in males and females, respectively (Table 2). This is in contrast with the meagre 0.1% of the total variance in susceptibility explained by all known coding mutations at RET³. Consequently, the MCS+9.7 enhancer mutation explains a 10–20-fold greater susceptibility variation than all other known RET mutations. However, our findings also show that a considerable number of additional loci may remain to be identified.

A final gender difference is that the mutant allele arises from mothers and fathers in 35 and 18 of the 53 informative families, respectively. This is significantly different from expectation ($P = 0.02$) and similar to the effect that we observed previously in linkage analysis of RET in a different series of families². The cause of this bias is unknown because RET is not known to be imprinted; however, whether RET shows specific imprinting in neuroblasts is unknown.

Discussion

A combination of human genetic, comparative genomic, functional, and population genetic analyses cumulatively demonstrates a common non-coding mutation at RET in HSCR. This common polymorphism is the explanation of our previous failure to find coding sequence mutations in the majority of HSCR cases, even in patients from families showing linkage to RET. This polymorphism means that many parents in multiplex families are homozygous for the variant allele and consequently fail to show segregation of RET. We conclude that RET mutations, coding and/or non-coding, are probably a necessary feature in all cases of HSCR. However, RET mutations are not sufficient for HSCR because the disease incidence also requires mutations at additional loci^{1,2,4}.

Our identification of the RET+3 mutation was aided by comparative sequence analysis and emphasized by its likely selection. This is not to suggest that additional RET variants cannot exist, but if they do not lie in identifiable functional sites then they cannot be confirmed. This finding has several implications for the genetic analyses of both mendelian and complex disease. First, mutation searches in human disease should include both the coding sequences of genes and neighbouring non-coding elements. This is critical not only to common complex disorders, in which non-coding mutations may conspire with mutations at additional genes for disease to occur, but also in rare mendelian phenotypes in which 10–15% of patients can have no recognized mutations despite incontrovertible evidence for a single known gene. Second, not all mutations for rare diseases are required to be rare or have 100% penetrance. Thus, the criterion of identifying mutations as sequence changes that are absent from controls may not be appropriate for a

significant fraction of alterations and may exclude legitimate mutations. Third, the inheritance patterns of single gene traits due to common variants are somewhat different from those that we have come to expect from rare mendelizing mutations, particularly when penetrance is not complete. Thus, apparent genetic heterogeneity in linkage or bilineal inheritance does not imply that mutations do not exist at a single locus.

One of the major challenges emphasized by our research is the critical need to define functional non-coding elements. A variety of non-coding elements are involved in transcription, translation, recombination, replication and repair, but we remain ignorant of the full nature and function of these sequences. Comparative genomics provides the current best avenue for recognizing such elements in a generic way, but this depends on the assumption that functional sites evolve recognizably more slowly than non-functional sites. These analyses have shown that only 1.5% of the human genome is devoted to coding exons, and as much as 3% to conserved non-coding sequences³⁹, implying that the latter might be particularly important as sites of mutation. Nevertheless, comparative methods are inefficient at recognizing function when it is determined by only a few nucleotides, such as at transcription-factor-binding sites. The recent focused efforts at experimental detection of non-coding regulatory elements are therefore very important to disease genetics²⁶. In the future, with a more complete knowledge of genomic function, mutation detection in human disease should encompass all functional elements, not only exons.

Our studies provide a molecular view to a multifactorial disorder: the most common mutation is non-coding, it has low (marginal) penetrance, the mutation has sex-dependent effects and explains only a small fraction of the total susceptibility to HSCR. Nevertheless, it has three features that are relevant to the analysis of common complex disorders. First, although the known protein-coding HSCR mutations have higher (51–72%) penetrance, their rarity in the population implies they explain only a minute fraction (0.1%) of the disorder. Additional genes or environmental factors are therefore necessary to explain disease incidence. Second, about 11% of our HSCR patients carry known RET coding mutations in addition to the RET+3:T variant. It is not unlikely that coding and non-coding mutations might act synergistically to affect disease penetrance; in other words, there might be more than one mutation per gene. Third, an enhancer mutation allows us to speculate that additional factors (proteins) interact with this element and can mitigate or attenuate its genetic effect on RET transcription. Thus, for common mutations, we expect that mutation penetrance will depend on other alleles and genes (genetic background), epigenetic effects (such as those associated with sex-linked gene dosage) or even the environment. □

Methods

Patient samples

We genotyped trios with 126 probands, all their parents (of whom 3 were affected) plus 24 unaffected siblings; for the penetrance studies we also genotyped additional probands for a total of 450 samples. All forms of HSCR (short segment, long segment, and total colonic aganglionosis) were represented in the patient sample. Of the ascertained cases, 11% presented with additional anomalies including defined neurocristopathies, chromosomal abnormalities (for example, trisomy 21) and other defects. Ascertainment was conducted under informed consent approved by the Institutional Review Board of the Johns Hopkins University School of Medicine. In addition to the HSCR patients and their families, we also genotyped 1,064 samples representing individuals from six continents from the CEPH Human Genome Diversity Panel (<http://www.cephb.fr/HGDP-CEPH-Panel/>; ref. 37).

SNP genotyping

We selected SNPs with a minimum minor allele frequency of 10%, with physical map locations covering the three genes RET, GALNACT-2 and RASGEF1A and emphasizing the associated region within RET⁸. From dbSNP (<http://www.ncbi.nlm.nih.gov/projects/SNP/>), we selected SNPs with known heterozygosity and/or SNPs with both alleles observed twice ('double hit' SNPs); we used markers for which robust genotyping assays could be developed. All SNPs are referred to by their rs numbers. Genotypes were generated with the fluorogenic 5' nuclease assay (Taqman; Applied Biosystems). A TECAN Genesis workstation was used for all liquid handling, thermal cycling was

Table 2 Genetic characteristics of the RET enhancer mutation

Genotype	Observed genotype counts†		Expected frequency‡	Penetrance ($\times 10^5$)§	
	Males	Females		Males	Females
CC	40	15	0.58	16.1 $\pm$ 2.2	6.2 $\pm$ 0.9
CT	50	17	0.37	34.5 $\pm$ 3.8	6.4 $\pm$ 1.3
TT	37	26	0.06	175.0 $\pm$ 22.9	35.9 $\pm$ 8.0
Risk ratio (γ)#				5.7	2.1
Variation (%)				2.63	1.14

The results assume a population incidence of 1 in 5,000, with S-HSCR and L-HSCR representing 80% and 20% of cases, respectively, and a sex ratio of 2:1 (M:F).

†Based on genotyping 185 affected individuals.

‡Estimated on the basis of allele frequencies of 0.76 (C) and 0.24 (T) on untransmitted chromosomes.

§Penetrance estimates are given with one standard deviation.

#Estimated by assuming a multiplicative model under which $\tau = \gamma/(1 + \gamma)$.

completed on MJ Research Tetrads, and end-point measurements were made on an ABI 7900 (Applied Biosystems). Genotypes were determined with SDS 2.1 (Applied Biosystems) and verified by the instrument operator. Of the samples, 10% ($n = 45$) were genotyped in duplicate for all 30 markers; no discrepancies were observed between the 1,350 paired replicate genotypes.

Transmission disequilibrium test

The TDT χ^2 -test statistic was used to identify significant deviation from the expected 1:1 mendelian transmission¹¹. The transmission frequency (τ) from heterozygous parents to offspring was estimated from all family genotype data at each SNP by maximum likelihood. We examined the following three hypotheses: single τ ; τ different by parent gender (τ_m, τ_f); or different transmission rates to male (b) and female (g) children (τ_b, τ_g). χ^2 tests with one degree of freedom based on the appropriate likelihood ratio were used to test whether $\tau = 1/2$, $\tau_m = \tau_f$ or $\tau_b = \tau_g$.

Haplotype reconstruction and EATDT

For family-based samples, haplotypes were inferred by using hap2, a method that combines traditional family-based reconstructions with population-based linkage disequilibrium information to achieve extremely accurate reconstruction within nuclear families¹². Haplotypes for control HGGP individuals were reconstructed with PHASE¹¹. EATDTs were performed, after haplotype reconstruction, for all sliding windows of all numbers of SNPs at all positions¹³. Within each window of any size, all observed haplotypes were tested for association by the TDT. To assess overall significance while accounting for multiple tests, 10^8 permutations were performed to estimate P .

Resequencing

Three resequencing experiments were performed and analysed to identify novel SNPs: first, DNA chip-based resequencing⁴² of the non-repeat sequence in a 90-kb interval containing *RET* in 32 Mennonites (15 HSCR cases and 17 controls); second, resequencing MCSs within *RET* intron 1 in 22 HSCR patients from families with *RET*-linkage but no identified coding sequence mutations; and third, resequencing 9 kb around *RET*+3 in four and eight individuals each homozygous for the *RET*+3:T and the *RET*+3:C allele, respectively. These analyses identified numerous rare and novel SNPs, additional low-frequency SNPs existing in dbSNP, and a high-frequency SNP within intron 1 enriched in patients, namely *RET*+3. In addition to *RET*+3, we identified variants within three additional intron 1 conserved elements (see below) by resequencing in HSCR patients.

Allele distribution at ENCODE loci

The ENCODE project²⁶ has identified all segregating sites at five loci on human chromosomes 2p16.3, 2q37.1, 4q26, 7q21.13 and 7q31.33, each about 500 kb in length. All SNPs were genotyped in the HapMap samples from four populations, namely Utah CEPH, Yoruba from Ibadan, Nigeria, Han Chinese from Beijing, and Japanese from Tokyo, Japan (www.HapMap.org). We estimated allele frequencies at 8,247 SNPs in the three continental regions (Europe, Africa and Asia; 60 independent individuals each) and compared them with those of the *RET*+3:T allele. We estimated the probability of observing the allele frequency as less than 5% in Yoruba, more than 25% in Europe and more than 40% in China/Japan in all 8,247 SNPs as 0.0046. To reduce effects of linkage disequilibrium, we sampled every second (4,121 SNPs), fourth (2,059 SNPs), eighth (1,028 SNPs) and sixteenth (512 SNPs) SNP to obtain probabilities of 0.0036, 0.0049, 0.0068 and 0.0059, respectively. An identical analysis with the F_{ST} statistics gave a P value of 0.027 (0.023–0.029).

Additional methods are provided in Supplementary Information.

Received 14 December 2004; accepted 15 February 2005; doi:10.1038/nature03467.

- Bolk, S. *et al.* A human model for multigenic inheritance: phenotypic expression in Hirschsprung disease requires both the *RET* gene and a new 9q31 locus. *Proc. Natl Acad. Sci. USA* **97**, 268–273 (2000).
- Gabriel, S. B. *et al.* Segregation at three loci explains familial and population risk in Hirschsprung disease. *Nature Genet.* **31**, 89–93 (2002).
- Chakravarti, A. & Lyonnet, S. in *The Metabolic and Molecular Bases of Inherited Disease* 8th edn (eds Scriver, C. R., Beaudet, A. R., Sly, W. & Valle, D.) Ch. 251, 6231–6255 (McGraw-Hill, New York, 2001).
- Carrasquillo, M. M. *et al.* Genome-wide association study and mouse model identify interaction between *RET* and *EDNRB* pathways in Hirschsprung disease. *Nature Genet.* **32**, 237–244 (2002).
- Borrego, S. *et al.* *RET* genotypes comprising specific haplotypes of polymorphic variants predispose to isolated Hirschsprung disease. *J. Med. Genet.* **37**, 572–578 (2000).
- Garcia-Barcelo, M. M. *et al.* Chinese patients with sporadic Hirschsprung's disease are predominantly represented by a single *RET* haplotype. *J. Med. Genet.* **40**, e122 (2003).
- Sancandi, M. *et al.* Single nucleotide polymorphic alleles in the 5' region of the *RET* proto-oncogene define a risk haplotype in Hirschsprung's disease. *J. Med. Genet.* **40**, 714–718 (2003).
- McCallion, A. S. *et al.* Genomic variation in multigenic traits: Hirschsprung disease. *Cold Spring Harb. Symp. Quant. Biol.* **68**, 373–381 (2003).
- Uyama, T. *et al.* Molecular cloning and expression of a second chondroitin N-acetylglucosaminyltransferase involved in the initiation and elongation of chondroitin/dermatan sulfate. *J. Biol. Chem.* **278**, 3072–3078 (2003).
- Sato, T. *et al.* Molecular cloning and characterization of a novel human β 1,4-N-acetylglucosaminyltransferase, β 4GalNAc-T3, responsible for the synthesis of N,N' -diacetylactosediamine, galNAc β 1-4GlcNAc. *J. Biol. Chem.* **278**, 47534–47544 (2003).
- Spielman, R. S., McGinnis, R. E. & Ewens, W. J. Transmission test for linkage disequilibrium: the insulin gene region and insulin-dependent diabetes mellitus (IDDM). *Am. J. Hum. Genet.* **52**, 506–516 (1993).
- Lin, S., Chakravarti, A. & Cutler, D. J. Haplotype and missing data inference in nuclear families. *Genome Res.* **14**, 1624–1632 (2004).
- Lin, S., Chakravarti, A. & Cutler, D. J. Exhaustive allelic transmission disequilibrium tests as a new approach to genome-wide association studies. *Nature Genet.* **36**, 1181–1188 (2004).
- Loots, G. G. *et al.* Identification of a coordinate regulator of interleukins 4, 13, and 5 by cross-species sequence comparisons. *Science* **288**, 136–140 (2000).
- Thomas, J. W. *et al.* Comparative analyses of multi-species sequences from targeted genomic regions. *Nature* **424**, 788–793 (2003).
- Kellis, M., Patterson, N., Endrizzi, M., Birren, B. & Lander, E. S. Sequencing and comparison of yeast species to identify genes and regulatory elements. *Nature* **423**, 241–254 (2003).
- Nobrega, M. & Pennacchio, L. A. Comparative genomic analysis as a tool for biological discovery. *J. Physiol.* **554**, 31–39 (2003).
- Bray, N., Dubchak, I. & Pachter, L. AVID: A global alignment program. *Genome Res.* **13**, 97–102 (2003).
- Margulies, E. H., Blanchette, M., Haussler, D. & Green, E. D. Identification and characterization of multi-species conserved sequences. *Genome Res.* **13**, 2507–2518 (2003).
- Shepherd, I. T., Pietsch, J., Elworthy, S., Kelsh, R. N. & Raible, D. W. Roles for GFRalpha1 receptors in zebrafish enteric nervous system development. *Development* **131**, 241–249 (2004).
- Shepherd, I. T., Beattie, C. E. & Raible, D. W. Functional analysis of zebrafish *GDNF*. *Dev. Biol.* **231**, 420–435 (2001).
- Rivas, E. & Eddy, S. R. Noncoding RNA gene detection using comparative sequence analysis. *BMC Bioinformatics* **2**, 8 (2001).
- Shoba, T., Dheen, S. T. & Tay, S. S. Retinoic acid influences the expression of the neuronal regulatory genes *Mash-1* and *c-ret* in the developing rat heart. *Neurosci. Lett.* **318**, 129–132 (2002).
- Batourina, E. *et al.* Vitamin A controls epithelial/mesenchymal interactions through *Ret* expression. *Nature Genet.* **27**, 74–78 (2001).
- Pitera, J. E., Smith, V. V., Woolf, A. S. & Milla, P. J. Embryonic gut anomalies in a mouse model of retinoic Acid-induced caudal regression syndrome: delayed gut looping, rudimentary cecum, and anorectal anomalies. *Am. J. Pathol.* **159**, 2321–2329 (2001).
- ENCODE Project Consortium. The ENCODE (ENCyclopedia Of DNA Elements) Project. *Science* **306**, 636–640 (2004).
- Haldane, J. B. S. The rate of mutation of human genes. *Heredity* **35** (suppl.), 267–273 (1948).
- Allison, A. C. G-6-PD deficiency in red blood cells of East Africans. *Nature* **186**, 531–532 (1960).
- Allison, A. C. & Clyde, D. F. Malaria in African children with deficient erythrocyte glucose-6-phosphate dehydrogenase. *Br. Med. J.* **5236**, 1346–1349 (1961).
- Motulsky, A. Metabolic polymorphisms and the role of infectious disease in human evolution. *Hum. Biol.* **32**, 28–62 (1960).
- Hill, A. V. *et al.* Common west African HLA antigens are associated with protection from severe malaria. *Nature* **352**, 595–600 (1991).
- Miller, L. H., Mason, S. J., Clyde, D. F. & McGinniss, M. H. The resistance factor to *Plasmodium vivax* in blacks. The Duffy-blood-group genotype, *FyFy*. *N. Engl. J. Med.* **295**, 302–304 (1976).
- Samson, M. *et al.* Resistance to HIV-1 infection in caucasian individuals bearing mutant alleles of the CCR-5 chemokine receptor gene. *Nature* **382**, 722–725 (1996).
- Dean, M. *et al.* Genetic restriction of HIV-1 infection and progression to AIDS by a deletion allele of the CCR5 structural gene. Hemophilia Growth and Development Study, Multicenter AIDS Cohort Study, Multicenter Hemophilia Cohort Study, San Francisco City Cohort, ALIVE Study. *Science* **273**, 1856–1862 (1996).
- Huang, Y. *et al.* The role of a mutant CCR5 allele in HIV-1 transmission and disease progression. *Nature Med.* **2**, 1240–1243 (1996).
- Collins, F. S. *et al.* New goals for the U.S. Human Genome Project: 1998–2003. *Science* **282**, 682–689 (1998).
- Lander, E. S. The new genomics: global views of biology. *Science* **274**, 536–539 (1996).
- Falconer, D. S. The inheritance of liability to diseases with variable age of onset, with particular reference to diabetes mellitus. *Ann. Hum. Genet.* **31**, 1–20 (1967).
- Waterston, R. H. *et al.* Initial sequencing and comparative analysis of the mouse genome. *Nature* **420**, 520–562 (2002).
- Cann, H. M. *et al.* A human genome diversity cell line panel. *Science* **296**, 261–262 (2002).
- Stephens, M., Smith, N. J. & Donnelly, P. A new statistical method for haplotype reconstruction from population data. *Am. J. Hum. Genet.* **68**, 978–989 (2001).
- Cutler, D. J. *et al.* High-throughput variation detection and genotyping using microarrays. *Genome Res.* **11**, 1913–1925 (2001).

Supplementary Information accompanies the paper on www.nature.com/nature.

Acknowledgements We thank members of the Chakravarti laboratory for their discussions on this manuscript, M. Kenton for assistance with family recruitment, and E. Margulies and M. Blanchette for help with multi-species sequence analysis. We thank the NISC Comparative Sequencing Program for generating the multi-species sequence data. We also acknowledge the many participants of the NISC Comparative Sequencing Program, especially the leadership provided by G. Bouffard and B. Blakesley. We also acknowledge the many participants of the NISC Comparative Sequencing Program, especially the leadership provided by G. Bouffard and B. Blakesley. This work was supported by grants from the US National Institute of Child Health and Development.

Competing interests statement The authors declare that they have no competing financial interests.

Correspondence and requests for materials should be addressed to A.C. (aravinda@jhmi.edu). GenBank accession numbers for genomic sequences reported in this paper are as follows: AC125509 and AC125512 (baboon), AC124166 (cat), AC138567 (chicken), RP43-171H18 (chimpanzee), AC124163 and AC124164 (cow), AC123973 (dog), AC124911 and AC125500 (fugu), AC122156 and AC124165 (pig), AC114881 (rat), AC135546 (tetra) and AC124155 (zebrafish).

DNA damage response as a candidate anti-cancer barrier in early human tumorigenesis

Jirina Bartkova¹, Zuzana Hořejší^{1,5}, Karen Koed², Alwin Krämer¹, Frederic Tort¹, Karsten Zieger², Per Guldberg¹, Maxwell Sehested³, Jahn M. Nesland⁴, Claudia Lukas¹, Torben Ørntoft², Jiri Lukas¹ & Jiri Bartek¹

¹Institute of Cancer Biology and Centre for Genotoxic Stress Research, Danish Cancer Society, Strandboulevarden 49, DK-2100 Copenhagen, Denmark

²Department of Clinical Biochemistry, Aarhus University Hospital, Skejby, DK-8200 Aarhus N, Denmark

³Department of Pathology, University Hospital, Frederik V's Vej 11, DK-2100 Copenhagen, Denmark

⁴Department of Pathology, The Norwegian Radium Hospital, University of Oslo, Ullernchaussen 70-0310 Oslo, Norway

⁵Institute of Molecular Genetics, Czech Academy of Sciences, Flemingovo nám. 2, Praha 6, CZ-16637, Czech Republic

During the evolution of cancer, the incipient tumour experiences 'oncogenic stress', which evokes a counter-response to eliminate such hazardous cells. However, the nature of this stress remains elusive, as does the inducible anti-cancer barrier that elicits growth arrest or cell death. Here we show that in clinical specimens from different stages of human tumours of the urinary bladder, breast, lung and colon, the early precursor lesions (but not normal tissues) commonly express markers of an activated DNA damage response. These include phosphorylated kinases ATM and Chk2, and phosphorylated histone H2AX and p53. Similar checkpoint responses were induced in cultured cells upon expression of different oncogenes that deregulate DNA replication. Together with genetic analyses, including a genome-wide assessment of allelic imbalances, our data indicate that early in tumorigenesis (before genomic instability and malignant conversion), human cells activate an ATR/ATM-regulated DNA damage response network that delays or prevents cancer. Mutations compromising this checkpoint, including defects in the ATM–Chk2–p53 pathway, might allow cell proliferation, survival, increased genomic instability and tumour progression.

Tumorigenesis is an evolutionary process that selects for genetic and epigenetic changes, allowing evasion of anti-proliferative and cell-death-inducing mechanisms that normally limit clonal expansion of somatic cells¹. Most tumours acquire genetic instability, but how early this occurs and whether it drives tumour development is unclear². Several mechanisms to constrain oncogenesis have been proposed, including hypoxia³, telomere attrition⁴ and induced expression of the Arf tumour suppressor (which is caused by the mitogenic overload experienced by incipient cancer cells)⁵. These are all conditions that can activate the tumour suppressor p53 (refs 1, 3–6). However, whether these mechanisms represent the major force(s) that guard against genetic instability and tumorigenesis is unknown. Recently, another possibility emerged from our observation⁷ that advanced carcinomas of the lung and breast show constitutive activation of Chk2, an effector kinase⁸ within the DNA damage network that is activated by the kinase ATM (Ataxia Telangiectasia Mutated) in response to DNA double-strand breaks^{9,10}. Furthermore, oncogenes such as Myc cause DNA damage in cultured cells^{11,12}. These findings led us to hypothesize that DNA damage checkpoints might become activated in the early stages of human tumorigenesis, leading to cell-cycle blockade or apoptosis and thereby constraining tumour progression.

DNA damage signalling in early bladder tumours

To determine whether Chk2 is activated in premalignant human tumours, we compared early, superficial lesions (stage Ta), early invasive (T1) and more advanced stages (T2–4) of urinary bladder cancer, all untreated by radiation or chemotherapy. Contrary to the negative staining of normal tissues, immunohistochemistry using a well-characterized antibody^{7,8} against activated Chk2 (phosphorylated at Thr 68) showed heterogeneous positive staining in the vast majority of the Ta lesions (Figs 1 and 2a). A similar pattern was seen in the T1 tumours, but the T2–4 carcinomas, although still commonly positive, showed moderately lower staining (Figs 1 and 2a).

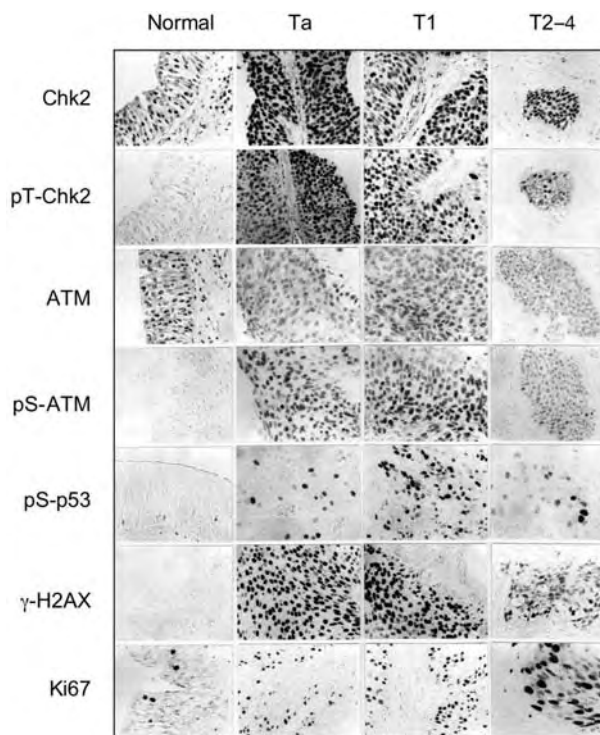


Figure 1 Constitutive activation of the ATM–Chk2–p53 pathway in human urinary bladder cancer. Immunohistochemistry of normal uroepithelium, early superficial lesions (Ta), earliest invasive (T1) and more advanced primary carcinomas (T2–4). Chk2 and ATM proteins are ubiquitously expressed, but Thr 68-phosphorylated Chk2 (pT-Chk2), Ser 1981-phosphorylated ATM (pS-ATM), Ser 15-phosphorylated p53 (pS-p53) and Ser 139-phosphorylated histone H2AX (γ-H2AX) are detectable only in tumour tissues. They are all present at the early stages of tumour development. Ki67 is a marker of proliferating cells. Original magnification, ×100.

These results support our hypothesis and raise the possibility that the upstream kinase ATM, which normally phosphorylates Chk2 (on Thr 68) in response to double-strand break-causing insults such as ionizing radiation^{9,10}, might be constitutively activated in tumours. Indeed, parallel tissue sections showed positive staining for Ser1981-phosphorylated ATM, a marker of ATM auto-activation¹³, as well as for the ATM substrates p53 (phosphorylated on Ser15) and histone H2AX (phosphorylated on Ser139, γ -H2AX)^{10,13,14} (Fig. 1 and Supplementary Figs S1, S2a).

DNA damage response precedes p53 mutations

Another prediction arising from the idea that DNA damage checkpoints act as a barrier against cancer and genetic instability (and a pressure selecting for p53 mutations), is that ATM–Chk2 pathway activation must precede the occurrence of pronounced genomic instability and p53 aberrations. To address these issues, we examined DNA isolated from 35 microdissected bladder tumours (stages Ta, T1, T2–4) for allelic imbalances, comparing single-nucleotide polymorphisms (SNPs) in blood and tumour DNA from each patient. This genome-wide estimate of genomic (in)stability was obtained using SNP arrays¹⁵, which included some 10,000 probes, of which about 2,600 were averagely heterozygous (informative). Reflecting the frequency of expected heterozygosity (present in blood DNA) versus aberrantly reduced heterozygosity in the tumour, we subdivided the tumours into groups according to their proportion of allelic imbalances (low, intermediate and high genomic instability; see Methods and Supplementary Fig. S2c). Levels of Chk2 and ATM phosphorylation (detected by immunohistochemistry) were already maximal in the most stable subset of tumours, and therefore preceded the development of pronounced genomic instability (Fig. 2b and Supplementary Fig. S2b).

We then estimated the relationship between DNA damage response activation and potential defects in the ATM–Chk2–p53

pathway using loss of heterozygosity at the respective gene loci, immunohistochemical analysis of protein abundance (Chk2 and ATM), and direct DNA sequencing (p53 and *CHK2*). These analyses revealed numerous abnormalities (Fig. 2c), including numerous p53 mutations and a previously unidentified splice mutation of *CHK2* that was accompanied by loss of heterozygosity and lack of the Chk2 protein (Fig. 2d). Notably, whereas only one aberration was detected among the tumours classified as stable according to the SNP array data, eight defects were found in the intermediate subset and 24 aberrations were identified in the unstable lesions (Fig. 2c). These results show that the ATM–Chk2–p53 cascade^{7–10} is activated in human bladder tumours before the occurrence of p53 mutations and/or defects in DNA damage signalling.

Constitutive DNA damage is shared by early human lesions

Analogous to bladder carcinomas, Thr68-phosphorylated Chk2 was detectable (at various levels) in the majority of 244 invasive carcinomas of the breast, colon or lung. This is in sharp contrast with negative samples from 87 normal, proliferating tissues (Fig. 3a, b and data not shown) and 26 inflammatory tissues (Supplementary Fig. S3). Furthermore, phosphorylated Chk2, ATM and H2AX were also found in the majority of pre-invasive carcinoma *in situ* (CIS, $n = 55$) lesions of the breast (Fig. 3a and data not shown), colorectal adenomas (Fig. 3b and see below) and lung hyperplasias (ref. 16 and data not shown). These data established that constitutive activation of the ATM–Chk2 pathway commonly occurs at pre-invasive stages of major types of human tumours.

Oncogenes induce DNA damage response in cultured cells

The above results raise a question about the identity of tumorigenic events that could evoke the observed DNA damage response in human cancer. First, we tested for the effects of cyclin E over-expression (Fig. 4a), an event common in carcinomas and reported to enhance genomic instability^{17–21}. In our U-2-OS-derived cells (a

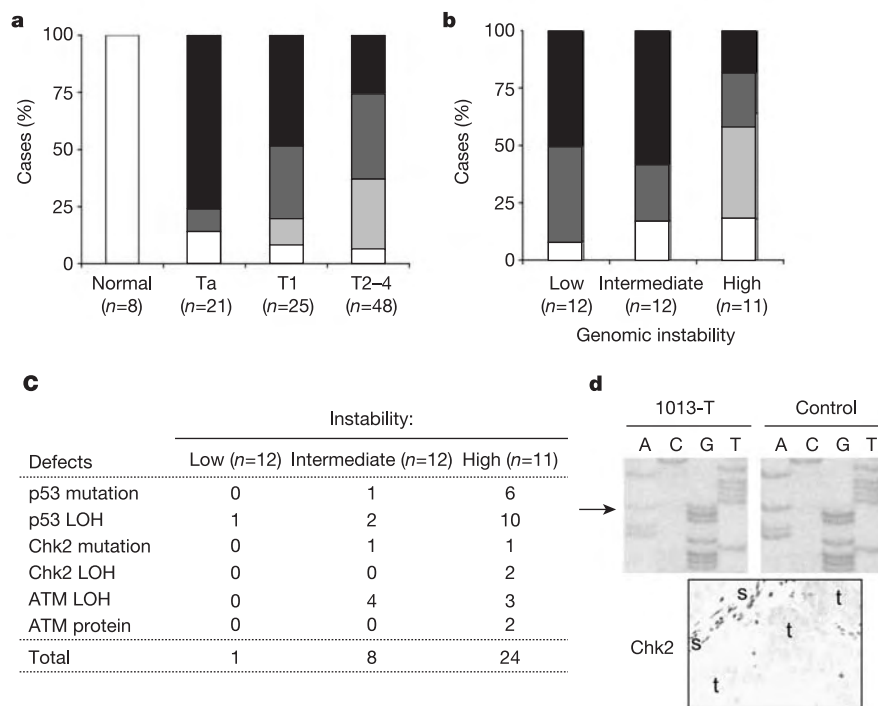


Figure 2 Chk2 activation precedes p53 mutations and genomic instability in bladder cancer. **a**, **b**, Summary graphs of Thr68-phosphorylated Chk2 (pT-Chk2) detected immunohistochemically in normal tissues and Ta, T1 and T2–4 lesions (**a**), and in lesions with different genomic instability (**b**). pT-Chk2 was either absent (white), or detected at low (light grey), medium (dark grey) or high (black) levels. **c**, Summary of tumour-

associated defects in the ATM–Chk2–p53 pathway in lesions with different genomic instability. LOH, loss of heterozygosity. **d**, Top panel, sequences of wild-type and mutant (1013-T) *CHK2* DNA. The arrow highlights an A to G substitution in tumour DNA. Bottom panel, lack of Chk2 detection (by immunohistochemistry) in tumour (t) but not in stromal cells (s) of the corresponding tissue sample.

checkpoint-proficient human osteosarcoma cell line with wild-type retinoblastoma (RB) and p53, cyclin E could be overexpressed in a tetracycline-repressible manner to reach levels comparable with the endogenous protein in breast cancer cell lines containing increased copies of the cyclin E gene (Fig. 4b). Biochemical analysis of cells with induced cyclin E expression showed time-dependent increases in Ser 15-phosphorylated p53, γ -H2AX, and Ser 966-phosphorylated cohesin SMC1 (Fig. 4a); these are all targets of the DNA damage response kinases ATM and ATR^{9,10,14,22,23}. Analogous to cyclin E, overexpression of tetracycline-regulatable Cdc25A²⁴, another proto-oncogene overexpressed in many carcinomas^{24,25}, evoked a DNA damage response (Fig. 4a). A similar response was also observed when oestrogen receptor-tagged E2F1 (ref. 26), an S-phase-promoting transcription factor commonly deregulated in cancer owing to aberrations of the RB pathway²⁷, was translocated into the nucleus upon addition of tamoxifen to the medium (Fig. 4a).

Consistent with our biochemical data (Fig. 4), the checkpoint-mediated phosphorylation markers were also detectable cytologically (Fig. 5a, b and Supplementary Fig. S4a, b), and were reminiscent of the heterogeneous staining patterns seen in the clinical specimens (Figs 1 and 3). Although the oncogene-induced phosphorylation of Chk2 (on Thr 68) and ATM (on Ser 1981) were detectable both biochemically and cytologically, they were less pronounced than γ -H2AX. This suggests that the ATR-H2AX/Chk1 pathway^{9,10,23} probably contributes to the overall response. Indeed, Chk1 was found to be phosphorylated on its activatory sites, Ser 317 and Ser 345 (Fig. 4c); of which Ser 345 is regarded as a preferential target for phosphorylation by ATR^{10,28,29}. Phosphorylated Chk1 was also detected by fluorescence microscopy two days after induction of cyclin E, Cdc25A or E2F1 (Fig. 5a). Another established target of ATR phosphorylation, the checkpoint protein Rad17 (refs 10, 30), was also phosphorylated (on Ser 645; Fig. 4b). Analogous to U-2-OS cells, ectopic expression of cyclin E induced a DNA damage response in human fibroblasts (Supplementary Fig. S5).

Our cell culture models show that S-phase-promoting oncogenes can activate the ATM/ATR-regulated DNA damage network.

Furthermore, unlike phosphorylations of p53 and H2AX induced by the Myc oncogene, which are a consequence of oxygen radical generation¹¹, the DNA damage response in our experiments was only marginally repressed by the antioxidant *N*-acetyl-L-cysteine (NAC; Fig. 4a and Supplementary Figs S4d, S5b, c). This indicates that the bulk of the DNA damage caused by the hyperproliferative oncogenic stimuli was independent of oxidative stress.

Because overexpressed cyclin E, Cdc25A and E2F1 share the ability to promote unscheduled S-phase entry^{26,31,32}, we argue that the DNA damage response could reflect deregulated DNA replication rather than oxidative stress. When cells experience replicational stress, they activate their replication checkpoint to delay S phase progression and G2/M transition³³. Consistent with this scenario, U-2-OS cells with induced cyclin E expression accumulated in S and G2 phases. The G2 arrest was particularly prominent by day 4 of time course experiments, and by day 6 many cells showed aberrantly enhanced DNA contents ($>4n$; Fig. 5c), probably reflecting partial re-replication of the genome uncoupled from cell division. The inability to enter mitosis correlated with massive induction of the Tyr 15-phosphorylated, inactive form of the key mitosis-promoting kinase Cdk1 (ref. 34; Fig. 5d), which is a marker of the effective checkpoint.

In further support of the idea that cyclin E overexpression alters replication dynamics, significantly longer replication tracks were detected using DNA fibre labelling³⁵, and increased amounts of the single-stranded-(ss)DNA-binding replication protein A (RPA)^{36,37} were bound to chromatin (Supplementary Fig. S4a, b). RPA loading on ssDNA occurs during normal replication initiation³⁸ and also on abnormal ssDNA intermediates that form during replicational stress^{36,39}; the latter scenario is consistent with the hyperphosphorylation of RPA³⁷ in our experiments. Importantly, RPA-coated ssDNA is a prerequisite (and serves as an activation signal) for the ATR-dependent checkpoint^{36,39}, and cyclin-induced aberrant DNA replication in yeast causes double-strand breaks^{40,41}, potentially deadly lesions that activate the ATM-Chk2 pathway in mammals^{9,10,33}.

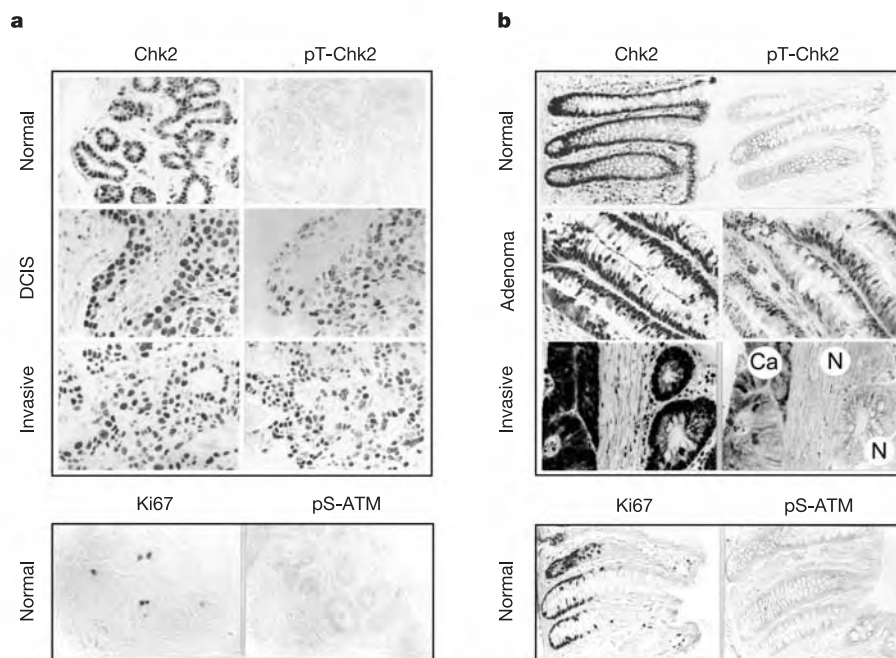


Figure 3 Phosphorylation and activation of Chk2 in early breast and colon tumours. **a**, Immunohistochemical staining of total Chk2 and Thr 68-phosphorylated Chk2 (pT-Chk2) in normal breast, early lesion of ductal carcinoma *in situ* (DCIS), and invasive ductal carcinoma. Bottom panel, normal breast shows limited cell proliferation (Ki67 marker) and no Ser 1981-phosphorylated ATM (pS-ATM). **b**, Total Chk2 and pT-Chk2 in

normal colon, precursor lesion of exophytic adenoma, and invasive carcinoma. Note pT-Chk2 staining in the carcinoma (Ca) but not adjacent normal epithelium (N). Bottom: normal colon crypts show high cell proliferation (Ki67) but no pS-ATM. Original magnification, $\times 100$ (**a**, **b**).

Oncogenes and checkpoint activation in early tumours *in vivo*

Together, our cell culture experiments suggest that tumorigenic abnormalities that deregulate DNA replication induce DNA damage and checkpoint activation. To validate this concept for early tumorigenesis *in vivo*, three predictions should be verified: (1) tumorigenic abnormalities that deregulate entry into S phase must occur with high frequency; (2) there should be evidence for replicational stress; and (3) activation of a functional checkpoint should be documented. We were able to verify each of these three predictions for premalignant lesions, using the analyses described below.

First, we found cases with RB defects among the early bladder lesions, and the majority of Ta lesions showed overexpressed cyclin E and/or unscheduled cyclin E expression (Fig. 6a, c). In addition, 64% of these lesions showed activating mutations in fibroblast growth factor receptor 3 (FGFR3) (ref. 42, Fig. 6a); these abnormalities might induce the DNA damage response in a manner analogous to effects of growth factor overload in human epithelial xenografts¹⁶. Finally, over half of the colon adenoma samples showed unscheduled or overexpressed cyclin E (Fig. 6a–c), and this correlated well with the DNA damage response (monitored using γ -H2AX as a marker; Supplementary Fig. S6a, b). Cyclin E abundance and its unscheduled expression during the cell cycle, defined as a ratio of cyclin E-positive to Ki67-positive cells (Fig. 6b), provides a readout for defective RB–E2F and/or Myc pathways, for amplification of the cyclin E gene, and for mutations in the

human *CDC4* gene (the product of which controls cyclin E turnover)^{20,21,26,27,32}. These results identify frequent aberrations of G1/S phase control in premalignant lesions.

Second, according to our proposal, double-strand breaks and loss of heterozygosity would probably occur at genomic sequences that are difficult to replicate: the fragile sites⁴³. Consistent with this, our SNP array data showed that among the genetically most stable early lesions, loss of heterozygosity at known fragile sites occurs 3–15 times more often than expected from random targeting (Fig. 6e). This contrasts with only a twofold preference for loss of heterozygosity at fragile sites among the more advanced and genetically unstable bladder tumours. The exact degree of preference should be interpreted with caution, given that some fragile sites harbour tumour suppressor genes and that their loss might provide a growth advantage. However, we note that fragile sites are targeted during the period in which we observe maximal DNA damage response, and that ATR is involved in the maintenance of fragile sites⁴³. Together with the emerging view that loss of heterozygosity at fragile sites is a ‘signature’ of stalled replication forks⁴³, these observations support our proposed concept.

Third, to show unequivocally that the activated checkpoint machinery affects its downstream cell-cycle targets, we examined samples of normal colon and adenomas for Tyr 15-phosphorylated Cdk1 (pY-Cdk1). In normal colon, pY-Cdk1 staining was predominantly cytoplasmic, detectable within the proliferative compart-

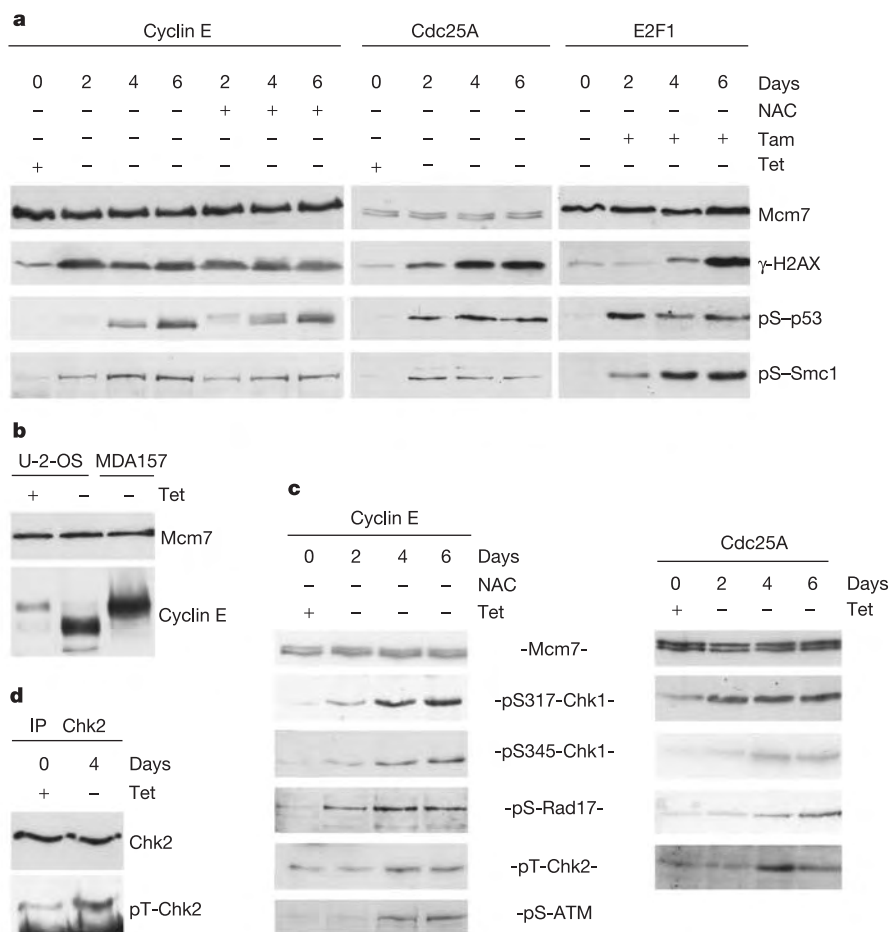


Figure 4 Overexpressed cyclin E, Cdc25A and E2F1 induce a DNA damage response in U-2-OS cells. **a**, Time course immunoblotting analyses using the indicated phospho-specific antibodies. NAC, antioxidant *N*-acetyl-L-cysteine; Tam, tamoxifen; Tet, tetracycline; Mcm7, loading control. **b**, Immunoblots showing cyclin E abundance after 2-day induction, compared with endogenous cyclin E levels in MDA157 cells with

amplified cyclin E gene. The altered mobility of ectopic cyclin E is explained in refs 31 and 32. **c**, Time course immunoblotting analyses using the indicated phospho-specific antibodies. **d**, Chk2 immunoprecipitates were analysed by immunoblotting for total Chk2 and pT-Chk2 before and 4 days after induction of cyclin E.

ment of the epithelial crypts in a modest fraction of cells (Fig. 6d). In contrast, over 50% of adenomas had enhanced pY-Cdk1, shown as higher staining intensity and larger groups or continuous sheets of positive cells (Fig. 6d). Staining of parallel sections revealed a good correlation between pY-Cdk1 and γ -H2AX (Fig. 6d and Supplementary Fig. S8a, c), consistent with the cell-cycle checkpoint being active in the same lesions (and the same areas within a lesion) that showed the DNA damage signalling. Reflecting their somewhat attenuated DNA damage response, invasive colorectal carcinomas showed less γ -H2AX and pY-Cdk1 staining than adenomas (Supplementary Fig. S8b). Finally, unlike in normal crypts where moderate pY-Cdk1 staining marked a subset of proliferating cells, the areas in adenomas strongly positive for pY-Cdk1 often showed only limited proliferation or were completely lacking in Ki67 staining (Supplementary Fig. S8c); this is consistent with prolonged cell-cycle arrest. These results indicate that the DNA damage response in precancerous lesions is functional and indeed affects the cell-cycle machinery.

Our data and those in ref. 16 suggest that tumorigenic events early in the progression of major human cancer types activate the ATR/ATM-regulated checkpoint through deregulated DNA replication and DNA damage, and thereby activate an inducible barrier against tumour progression and genetic instability (Fig. 6f). We

propose that candidate inducers of such a response include aberrations that grossly overstimulate growth factor pathways, affect RB or impair DNA replication downstream of RB (such as deregulated E2F1 (refs 25–27, 44) and overexpressed cyclin E^{17–21} or Cdc25A^{24,25}), and mutations in DNA repair genes such as *BRCA2* (ref. 45). The observed DNA damage might reflect abnormalities in pre-replication complex maturation and/or stalled or collapsed replication forks (which are known inducers of the ATR–H2AX/Chk1 cascade^{9,23,33,36,39}), followed by the generation of double-strand breaks^{40,41,45} and consequent activation of the ATM–H2AX/Chk2–p53 pathway^{8–10}. We speculate that such responses might limit the progression of lesions, possibly contributing to long ‘latency’ periods or failure of early lesions to ever become malignant. Tumour progression under such circumstances probably relies on selection of cells defective in their DNA damage response components (such as ATM or p53), with compromised cell-cycle arrest, senescence and apoptosis. Oncogene-induced DNA damage, particularly double-strand breaks, might also enhance genomic instability, especially when checkpoint responses have been evaded.

Whether the DNA damage response reported here represents the predominant mode for constraining tumorigenesis remains to be established. However, it is apparent that p53 induction by the Arf tumour suppressor⁵ cannot account for our data, because the DNA

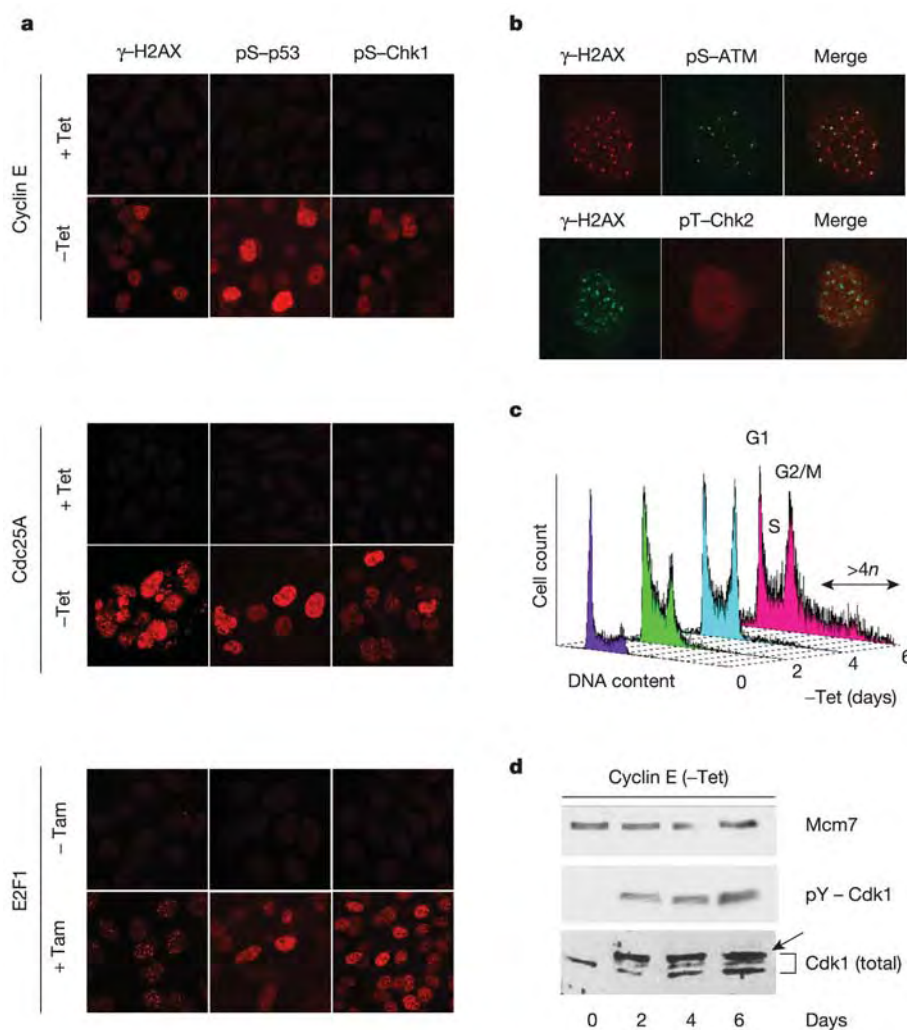


Figure 5 DNA damage response to overexpressed oncogenes in U-2-OS cells.

a, Confocal immunofluorescence microscopy images of cells stained for γ -H2AX, pS-p53 or pS-Chk1, after 4 days induction of cyclin E or Cdc25A (induced by removing tetracycline from the medium, -Tet), or induction of E2F1 (induced by tamoxifen (+Tam), causing translocation to the nucleus). **b**, Double-immunofluorescence colocalization of

γ -H2AX and pS-ATM foci (top), and pT-Chk2 staining in γ -H2AX-foci-positive nuclei (bottom) in cells with induced cyclin E. **c**, Time course analysis of flow cytometry. **d**, Time course immunoblotting analysis of Mcm7 (loading marker), pY-Cdk1 and total Cdk1. Arrow shows the Tyr 15-phosphorylated form of Cdk1.

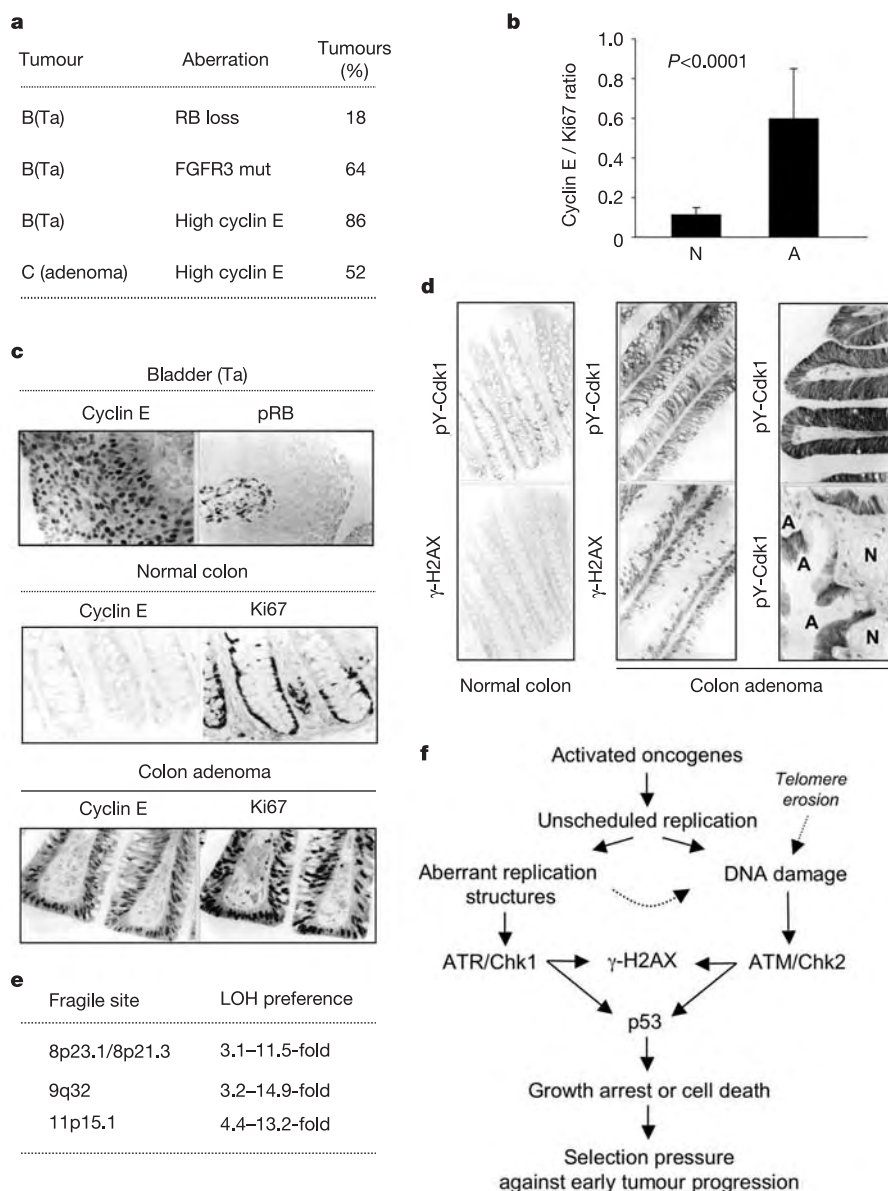


Figure 6 Oncogenes and DNA damage response in early lesions. **a**, Tumorigenic aberrations in bladder (B) and colon (C) lesions. Frequent FGFR3 mutations ($n = 33$) were R248C, S249C and Y375C. High cyclin E levels denote a cyclin E/Ki67-positive cell ratio over 0.5. **b**, Cyclin E/Ki67 ratios in normal colon (N; $n = 30$) and adenomas (A; $n = 29$). Data show means $\pm$ s.d. **c**, Immunohistochemistry showing cyclin E overexpression and loss of RB protein (pRB) (top panel), and normal (middle panel) and high (bottom panel)

cyclin E/Ki67 ratios. **d**, Immunohistochemical staining of Tyr 15-phosphorylated Cdk1 (pY-Cdk1) and γ -H2AX in colon adenoma (A) and normal epithelium (N). **e**, Loss of heterozygosity (LOH) at fragile sites occurs preferentially (at higher than the expected (random) frequency) in the genetically most stable early bladder lesions. **f**, Model of the DNA damage response in tumorigenesis.

damage response does not rely on Arf¹. On the other hand, telomere dysfunction, a phenomenon that might occur in premalignant tumours^{46,47}, can mimic DNA damage and activate the ATM/ATR-dependent checkpoints, thereby contributing to replicative senescence in cultured cells^{48,49}. We propose that oncogene-induced DNA damage and dysfunctional telomeres might converge to activate the DNA damage response, acting as a common mechanism to prevent progression of preneoplastic lesions (Fig. 6f). From a broader perspective, the DNA damage response machinery might play an important role in fundamental biological processes such as ageing and tumorigenesis. □

Methods

Antibodies

We used antibodies against total and Thr 68-phosphorylated Chk2 (refs 7 and 8), cyclin E (Novocastra Laboratories), cyclin A (Novocastra Laboratories), γ -H2AX (Upstate), RPA

p32 (Neomarkers), Ser 1981-phosphorylated ATM (Rockland or from C. Bakkenist and M. B. Kastan), ATM (from Y. Shiloh), Ki67 (Dako), γ -H2AX, Ser 15-phosphorylated p53, Ser 317-phosphorylated Chk1, Ser 345-phosphorylated Chk1 and Ser 645-phosphorylated Rad17 (all from Cell signalling), and Tyr 15-phosphorylated Cdk1 (Calbiochem) and Ser 966-phosphorylated Smc1 (Abcam).

Immunohistochemistry

Tumours and control tissues were obtained from the tissue banks of four institutes in Denmark and Norway, with the approval of local ethical committees. The early lesions from bladder (Ta), breast (CIS) and lung (hyperplasias) showed broad spectra of histopathological grading, whereas the majority of colon adenomas were of high grade. Indirect immunoperoxidase staining on formaldehyde-fixed, de-paraffinized tissue sections or on formalin-fixed cultured cells was performed as described⁷, using the Vectastain Elite kit and nickel sulphate enhancement without nuclear counterstaining. The immunostaining patterns were evaluated by an experienced pathologist and scored as (1) negative (no positive staining or up to 1% of scattered positive cells); (2) low (heterogeneous staining, where an area corresponding to at least 20% of the section showed 2–10% positive cells); (3) medium (heterogeneous, with at least 20% of the section showing 10–50% positive cells); or (4) high positivity (variable to almost homogeneous

staining, with at least 20% of the section showing 51–90% positive cells).

Cell culture

Cells were grown in DMEM medium supplemented with 10% fetal calf serum and antibiotics. Construction, characterization and induction of U-2-OS-derived cell lines with tetracycline-repressible expression of human Cdc25A²⁴ and tamoxifen-inducible nuclear translocation and activity of human E2F1 (ref. 26) were performed as described. U-2-OS-derived cells with tetracycline-repressible expression of wild-type, full-length human cyclin E were constructed using previously reported methods and vectors^{24,32}. Near-homogeneous expression of de-repressed cyclin E was seen in 90–95% of the cells by immunofluorescence, and despite the slightly altered gel mobility (owing to splicing also noticed in other regulatable systems³¹), the ectopic cyclin E was functional as judged from its selective binding to Cdk2 and stimulation of its kinase activity^{31,32} (data not shown). Retrovirus-mediated expression of cyclin E in normal human diploid fibroblasts was achieved as described¹⁸.

Biochemical analyses

Procedures for gel electrophoresis, immunoblotting and immunoprecipitation followed published protocols^{8,32}. For immunoprecipitation of total Chk2, we used 800 µg of cellular protein lysate and 10 µl of rabbit primary antibody (H-300:sc-9064, Santa Cruz).

Immunofluorescence

The fixation, permeabilization and immunofluorescence staining procedures, as well as the details of laser confocal microscopy, have been previously published⁸.

Mutation analysis

Genomic DNA from laser-microdissected tumour tissue or blood cells was examined for FGFR3 mutations by direct sequencing. Chk2 and p53 mutations were assessed using denaturing gradient gel electrophoresis of polymerase chain reaction (PCR)-amplified fragments that covered the entire coding sequence and all splice sites; any aberrantly migrating fragments were sequenced. The primers used for amplification will be provided upon request. The identified mutations in p53 included missense mutations and one mutation resulting in a stop codon downstream of the single-nucleotide deletion (C at position 482). Loss of heterozygosity was assessed using the SNP array data, on the basis of allelic imbalances detected by SNP probes within the regions of the Chk2, p53 and ATM genes. Both the SNPs and the genes were positioned according to the same build of the human genome sequence (the UCSC April 2003 assembly, version hg15, which is based on NCBI Build 33).

SNP array analysis

Tumours for DNA extraction were frozen immediately after surgery and stored at –80 °C. Before DNA extraction, the tumours were transferred to Tissue-Tek (Sakura Finetek) and tumour tissue was crudely microdissected from tumour sections (20-µm thick). Control DNA was purified from blood of the same patient. DNA was extracted from tumour tissue and blood using a PUREGENE protocol (Gentra SYSTEMS), according to the manufacturer's instructions. The DNA concentration was determined using a spectrophotometer, and diluted to a concentration of 50 ng µl⁻¹. The Single Primer Assay Protocol for GeneChip Mapping 10K Early Access Array was performed according to the manufacturer's instructions (Affymetrix). We developed an algorithm based on a Hidden Markov Model to score allelic imbalances³⁰. The numbers of SNPs heterozygous in blood DNA were compared with corresponding tumour DNA, and the tumours were divided into 3 groups as follows: high instability (from 1270–1720 heterozygous SNPs preserved in the tumour; $n = 11$); intermediate instability (1900–2360; $n = 12$) and low instability (2390–2603 heterozygous SNPs; $n = 12$). The frequency of loss of heterozygosity at fragile sites compared with other genomic sites was corrected for the number of SNPs in the fragile sites as a fraction of the total number of SNPs on the array.

Received 15 December 2004; accepted 18 February 2005; doi:10.1038/nature03482.

- Lowe, S. W., Cepero, E. & Evan, G. I. Intrinsic tumour suppression. *Nature* **432**, 307–315 (2004).
- Lengauer, C., Kinzler, K. W. & Vogelstein, B. Genetic instabilities in human cancers. *Nature* **396**, 643–649 (1998).
- Graeber, T. G. *et al.* Hypoxia induces accumulation of p53 protein, but activation of a G1-phase checkpoint by low-oxygen conditions is independent of p53 status. *Mol. Cell. Biol.* **14**, 6264–6277 (1994).
- Counter, C. M. *et al.* Telomere shortening associated with chromosome instability is arrested in immortal cells which express telomerase activity. *EMBO J.* **11**, 1921–1929 (1992).
- Zindy, F. *et al.* Arf tumour suppressor promoter monitors latent oncogenic signals in vivo. *Proc. Natl Acad. Sci. USA* **100**, 15930–15935 (2003).
- Vogelstein, B., Lane, D. & Levine, A. J. Surfing the p53 network. *Nature* **408**, 307–310 (2000).
- DiTullio, R. A. Jr *et al.* 53BP1 functions in an ATM-dependent checkpoint pathway that is constitutively activated in human cancer. *Nature Cell Biol.* **4**, 998–1002 (2002).
- Lukas, C., Falck, J., Bartkova, J., Bartek, J. & Lukas, J. Distinct spatiotemporal dynamics of mammalian checkpoint regulators induced by DNA damage. *Nature Cell Biol.* **5**, 255–260 (2003).
- Kastan, M. B. & Bartek, J. Cell-cycle checkpoints and cancer. *Nature* **432**, 316–323 (2004).
- Shiloh, Y. ATM and related protein kinases: safeguarding genome integrity. *Nature Rev. Cancer* **3**, 155–168 (2003).
- Vafa, O. *et al.* c-Myc can induce DNA damage, increase reactive oxygen species, and mitigate p53 function: a mechanism for oncogene-induced genetic instability. *Mol. Cell* **9**, 1031–1044 (2002).
- Lindström, M. S. & Wiman, K. G. Myc and E2F1 induce p53 through p14ARF-independent mechanisms in human fibroblasts. *Oncogene* **22**, 4993–5005 (2003).
- Bakkenist, C. J. & Kastan, M. B. DNA damage activates ATM through intermolecular autophosphorylation and dimer dissociation. *Nature* **421**, 499–506 (2003).
- Rogakou, E. P., Pilch, D. R., Orr, A. H., Ivanova, V. S. & Bonner, W. M. DNA double-stranded breaks

- induce histone H2AX phosphorylation on serine 139. *J. Biol. Chem.* **273**, 5858–5868 (1998).
- Lindblad-Toh, K. *et al.* Loss-of-heterozygosity analysis of small-cell lung carcinomas using single-nucleotide polymorphism arrays. *Nature Biotechnol.* **18**, 1001–1005 (2000).
- Gorgoulis, V. G. *et al.* Activation of the DNA damage checkpoint and genomic instability in human precancerous lesions. *Nature* doi:10.1038/nature03485 (this issue).
- Spruck, C. H., Won, K. A. & Reed, S. I. Deregulated cyclin E induces chromosome instability. *Nature* **401**, 297–300 (1999).
- Minella, A. C. *et al.* p53 and p21 form an inducible barrier that protects cells against cyclin E-cdk2 deregulation. *Curr. Biol.* **12**, 1817–1827 (2002).
- Eskholm-Reed, S. *et al.* Deregulation of cyclin E in human cells interferes with prereplication complex assembly. *J. Cell Biol.* **165**, 789–800 (2004).
- Reed, S. E. *et al.* Mutation of hCDC4 leads to cell cycle deregulation of cyclin E in cancer. *Cancer Res.* **64**, 795–800 (2004).
- Rajagopalan, H. *et al.* Inactivation of hCDC4 can cause chromosomal instability. *Nature* **428**, 77–81 (2004).
- Kim, S. T., Xu, B. & Kastan, M. B. Involvement of the cohesin protein, Smc1, in Atm-dependent and independent responses to DNA damage. *Genes Dev.* **16**, 560–570 (2002).
- Ward, I. M. & Chen, J. Histone H2AX is phosphorylated in an ATR-dependent manner in response to replicational stress. *J. Biol. Chem.* **276**, 47759–47762 (2001).
- Mailand, N. *et al.* Rapid destruction of human Cdc25A in response to DNA damage. *Science* **288**, 1425–1429 (2000).
- Cangi, M. G. *et al.* Role of the Cdc25A phosphatase in human breast cancer. *J. Clin. Invest.* **106**, 753–761 (2000).
- Müller, H. *et al.* E2Fs regulate the expression of genes involved in differentiation, development, proliferation, and apoptosis. *Genes Dev.* **15**, 267–285 (2001).
- Bartek, J., Bartkova, J. & Lukas, J. The retinoblastoma protein pathway in cell cycle control and cancer. *Exp. Cell Res.* **237**, 1–6 (1997).
- Zhao, H. & Piwnicka-Worms, H. ATR-mediated checkpoint pathways regulate phosphorylation and activation of human Chk1. *Mol. Cell. Biol.* **21**, 4129–4139 (2001).
- Gatei, M. *et al.* ATM and NBS1-dependent phosphorylation of Chk1 on Ser-317 in response to ionizing radiation. *J. Biol. Chem.* **278**, 14806–14811 (2003).
- Bao, S. *et al.* ATR/ATM-mediated phosphorylation of human Rad17 is required for genotoxic stress responses. *Nature* **411**, 969–974 (2001).
- Resnitzky, D., Gossen, M., Bujard, H. & Reed, S. I. Acceleration of the G1/S phase transition by expression of cyclins D1 and E with an inducible system. *Mol. Cell. Biol.* **14**, 1669–1679 (1994).
- Santoni-Rugiu, E., Falck, J., Mailand, N., Bartek, J. & Lukas, J. Involvement of Myc activity in a G1/S-promoting mechanism parallel to the pRB/E2F pathway. *Mol. Cell. Biol.* **20**, 3497–3509 (2000).
- Bartek, J., Lukas, C. & Lukas, J. Checking on DNA damage in S phase. *Nature Rev. Mol. Cell Biol.* **5**, 792–804 (2004).
- Lew, D. J. & Kornbluth, S. Regulatory roles of cyclin dependent kinase phosphorylation in cell cycle control. *Curr. Opin. Cell Biol.* **8**, 795–804 (1996).
- Merrick, C. J., Jackson, D. & Diffley, J. F. X. Visualization of altered replication dynamics after DNA damage in human cells. *J. Biol. Chem.* **279**, 20067–20075 (2004).
- Zou, L. & Elledge, S. J. Sensing DNA damage through ATRIP recognition of RPA-ssDNA complexes. *Science* **300**, 1542–1548 (2003).
- Binz, S. K., Sheehan, A. M. & Wold, M. S. Replication protein A phosphorylation and the cellular response to DNA damage. *DNA Repair (Amst.)* **3**, 1015–1024 (2004).
- Walter, J. & Newport, J. Initiation of eukaryotic DNA replication: origin unwinding and sequential chromatin association of Cdc45, RPA and DNA polymerase α . *Mol. Cell* **5**, 617–627 (2000).
- Osborn, A. J., Elledge, S. J. & Zou, L. Checking on the fork: the DNA-replication stress-response pathway. *Trends Cell Biol.* **12**, 509–516 (2002).
- Tanaka, S. & Diffley, J. F. Deregulated G1-cyclin expression induces genomic instability by preventing efficient pre-RC formation. *Genes Dev.* **16**, 2639–2649 (2002).
- Lengronne, A. & Schwob, E. The yeast CDK inhibitor Sic1 prevents genomic instability by promoting replication origin licensing in late G1. *Mol. Cell* **9**, 1067–1078 (2002).
- Hart, K. C. *et al.* Transformation and Stat activation by derivatives of FGFR1, FGFR3, and FGFR4. *Oncogene* **19**, 3309–3320 (2000).
- Casper, A. M., Nghiem, P., Arlt, M. F. & Glover, T. W. ATR regulates fragile site stability. *Cell* **111**, 779–789 (2002).
- Rogoff, H. A. *et al.* Apoptosis associated with deregulated E2F activity is dependent on E2F1 and Atm/Nbs1/Chk2. *Mol. Cell. Biol.* **24**, 2968–2977 (2004).
- Lomonosov, M., Anand, S., Sangrithi, M., Davies, R. & Venkataraman, A. R. Stabilization of stalled DNA replication forks by the BRCA2 breast cancer susceptibility protein. *Genes Dev.* **17**, 3017–3022 (2003).
- Meeker, A. K. *et al.* Telomere length abnormalities occur early in the initiation of epithelial carcinogenesis. *Clin. Cancer Res.* **10**, 3317–3326 (2004).
- Chin, K. *et al.* *In situ* analyses of genome instability in breast cancer. *Nature Genet.* **36**, 984–988 (2004).
- d'Adda di Fagnana, F. *et al.* DNA damage checkpoint response in telomere-initiated senescence. *Nature* **426**, 194–198 (2003).
- Takai, H., Smogorzewska, A. & de Lange, T. DNA damage foci at dysfunctional telomeres. *Curr. Biol.* **13**, 1549–1556 (2003).
- Eddy, S. R. Profile hidden Markov models. *Bioinformatics* **14**, 755–763 (1998).

Supplementary Information accompanies the paper on www.nature.com/nature.

Acknowledgements We thank Y. Shiloh, C. Bakkenist, M. Kastan, M. Welcker, B. Clurman, V. Gorgoulis and K. Helin for reagents; M.H. Lee, D. Lützholtz, A. Arnt Kjerulf and L.-L. Christensen for technical assistance; and C. Wiuf for data analysis. This work was supported by the Danish Cancer Society, the Alfred Benzon Foundation and the European Commission.

Competing interests statement The authors declare that they have no competing financial interests.

Correspondence and requests for materials should be addressed to J.B. (jb@cancer.dk).

Nucleosynthetic signatures of the first stars

Anna Frebel¹, Wako Aoki², Norbert Christlieb^{2,3}, Hiroyasu Ando², Martin Asplund¹, Paul S. Barklem⁴, Timothy C. Beers⁵, Kjell Eriksson⁴, Cora Fechner³, Masayuki Y. Fujimoto⁶, Satoshi Honda², Toshitaka Kajino², Takeo Minezaki⁷, Ken'ichi Nomoto⁸, John E. Norris¹, Sean G. Ryan⁹, Masahide Takada-Hidai¹⁰, Stelios Tsangarides⁹ & Yuzuru Yoshii⁷

¹Research School of Astronomy & Astrophysics, The Australian National University, Cotter Road, Weston Creek, Australian Capital Territory 2611, Australia

²National Astronomical Observatory of Japan, 2-21-1, Osawa, Mitaka, Tokyo 181-8588, Japan

³Hamburger Sternwarte, Gojenbergsweg 112, 21029 Hamburg, Germany

⁴Department of Physics and Space Sciences, Uppsala Astronomical Observatory, Box 515, SE-751 20 Uppsala, Sweden

⁵Department of Physics and Astronomy, and Joint Institute for Nuclear Astrophysics (JINA), Michigan State University, East Lansing, Michigan 48824, USA

⁶Department of Physics, Hokkaido University, Sapporo, Hokkaido 060-0810, Japan

⁷Institute of Astronomy, School of Science, University of Tokyo, Mitaka, Tokyo 181-0015, Japan

⁸Department of Astronomy, School of Science, University of Tokyo, Tokyo 113-0033, Japan

⁹Department of Physics and Astronomy, Open University, Walton Hall, Milton Keynes MK7 6AA, UK

¹⁰Liberal Arts Education Center, Tokai University, 1117 Kitakaname, Hiratsuka-shi, Kanagawa 259-1292, Japan

The chemically most primitive stars provide constraints on the nature of the first stellar objects that formed in the Universe; elements other than hydrogen, helium and traces of lithium present within these objects were generated by nucleosynthesis in the very first stars. The relative abundances of elements in the surviving primitive stars reflect the masses of the first stars, because the pathways of nucleosynthesis are quite sensitive to stellar masses. Several models^{1–5} have been suggested to explain the origin of the abundance pattern of the giant star HE0107–5240, which hitherto exhibited the highest deficiency of heavy elements known^{1,6}. Here we report the discovery of HE1327–2326, a subgiant or main-sequence star with an iron abundance about a factor of two lower than that of HE0107–5240. Both stars show extreme overabundances of carbon and nitrogen with respect to iron, suggesting a similar origin of the abundance patterns. The unexpectedly low Li and high Sr abundances of HE1327–2326, however, challenge existing theoretical understanding: no model predicts the high Sr abundance or provides a Li depletion mechanism consistent with data available for the most metal-poor stars.

The star HE1327–2326 (coordinates right ascension, RA = 13 h 30 min 06 s and declination, $\delta = -23^\circ 41' 51''$ at Equinox 2000; apparent visual magnitude $V = 13.5$ mag) was found among 1,777 bright ($10 < B < 14$) and apparently metal-poor objects with partly saturated spectra selected from the Hamburg/ESO objective-prism survey (HES)⁷. Its extreme metal-weakness was first appreciated in a medium-resolution ($\delta\lambda \approx 0.5$ nm) follow-up spectrum obtained in April 2003 at the 3.6-m telescope of the European Southern Observatory, Chile. In May and June 2004, observations with the High Dispersion Spectrograph⁸ at the Subaru telescope, Hawaii, were taken. Figure 1 shows a portion of this spectrum and gives more details on the discovery.

An important difference between HE0107–5240 and HE1327–2326 is their evolutionary status. HE0107–5240 is a giant and has evolved off the main-sequence branch and up the red-giant branch.

Consequently, internal mixing has possibly dredged up processed material (such as C, N) from the stellar interior to the surface and altered the abundances of this object. In contrast, HE1327–2326 is relatively unevolved and located on either the main-sequence or the subgiant branch.

We use model atmosphere techniques and the assumption of local thermodynamical equilibrium (LTE) to perform our chemical abundance analysis with non-LTE corrections where available. There are no significant differences between the subgiant and the main-sequence solution. See Table 1 for more details on the analysis procedure and the derived abundances.

HE1327–2326 sets a new record for the star with the lowest iron abundance: $[\text{Fe}/\text{H}]_{\text{non-LTE}} = -5.4 \pm 0.2$ (where $[A/B] = \log_{10}(N_A/N_B) - \log_{10}(N_A/N_B)_\odot$ for the number N of atoms of elements A and B and subscript $\odot$ refers to the Sun), derived from four detected weak Fe I lines and corrected for non-LTE. This low value corresponds to $\sim 1/250,000$ of the solar Fe value. To illustrate the nature of such extremely Fe-deficient stars, in Fig. 2 we compare our

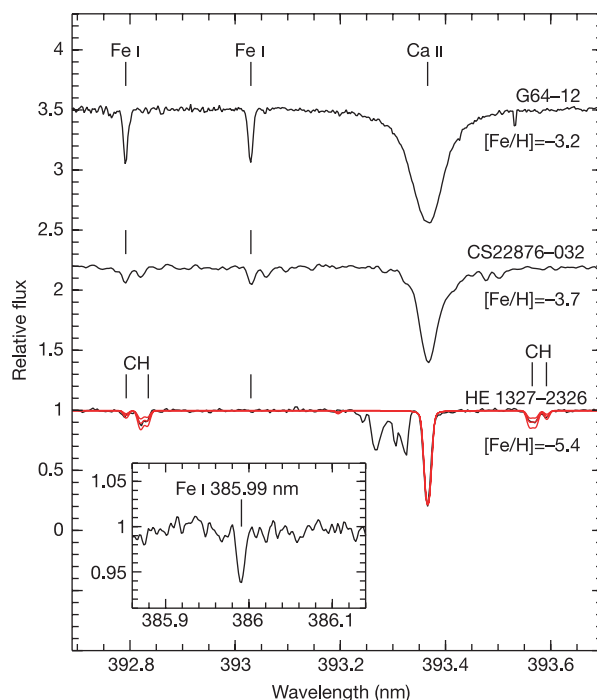


Figure 1 Comparison of high-resolution spectra of HE1327–2326 with G64–12 and CS22876–032. The latter is a double-lined spectroscopic binary. All three stars have a similar effective temperature and gravity. In HE1327–2326 we note the absence of the Fe I line at 393 nm together with the appearance of CH lines. Our best fit for the Ca II K ($\lambda = 393.4$ nm) and CH band abundances is plotted in red. Additional CH fits (in red) for our carbon abundance $[\text{C}/\text{Fe}] = 4.1 \pm 0.2$ dex are shown as well. The inset shows our strongest detected Fe I line (0.64 pm equivalent width) at 386 nm in the Subaru data from which we derive $[\text{Fe}/\text{H}]_{\text{non-LTE}} = -5.4 \pm 0.2$. From our medium-resolution spectrum we initially estimated $[\text{Fe}/\text{H}] = -4.0$ for HE1327–2326 using the apparent strength of the Ca II K line index KP of Beers *et al.*²¹ and an approximate colour. The high-resolution data revealed the presence of strong interstellar Ca II K that had not been resolved at the lower resolution. See the prominent feature blueward of the Ca II K line. The interstellar Ca is consistent with a colour excess of $E(B - V) = 0.08$ along the line of sight to HE1327–2326. The presence of CH lines near the Ca II K line in HE0107–5240 caused a similar problem in the initial recognition of its extreme metal-deficiency. These effects should be taken into account in future searches for more stars with similar heavy-element deficiency. Our high-resolution (resolving power $R = \lambda/\Delta\lambda = 50,000$), high signal-to-noise ratio ($S/N = 160$ per pixel at 400 nm) data cover the wavelength range of 304 to 674 nm.

abundances with those of HE0107–5240 ($[\text{Fe}/\text{H}]_{\text{non-LTE}} = -5.2$; refs 1, 6 and 9), the only other star known to have $[\text{Fe}/\text{H}] < -4.0$. Both objects have very large overabundances of C and N relative to Fe. Although both stars also share a deficiency of Ca and Ti as well as Fe, the Mg and Al abundances relative to Fe in HE1327–2326 are more than one order of magnitude higher than those of HE0107–5240.

Surprisingly, we do not detect the Li I doublet at 670.7 nm in HE1327–2326, although it is found in similarly unevolved metal-poor stars that permit observational determination of the primordial Li abundance and the associated baryon density of the Universe. Our upper limit for Li of $A(\text{Li}) = 1.6$ (where $A(\text{Li}) = \log_{10}(N(\text{Li})/N(\text{H})) + 12.00$) is significantly lower than the primordial Li value of $A(\text{Li}) = 2.62 \pm 0.05$, inferred from the baryon density estimated by the Wilkinson Microwave Anisotropy Probe together with standard Big Bang nucleosynthesis¹⁰. It is even below $A(\text{Li}) \cong 2.0$, found by Ryan *et al.*¹¹ at $[\text{Fe}/\text{H}] = -3.5$. This apparent absence does not result from the evolutionary status of HE1327–2326, because in more evolved metal-poor subgiants such as HD140283, Li has not been depleted. Li weakness occurs rarely in metal-poor stars near the main-sequence turnoff: at effective temperature $T_{\text{eff}} \approx 6,200$ K only 1 in ~ 20 shows this effect¹¹. In HE1327–2326, the weak lines show no evidence for rotational broadening and offer no support for the ultra-Li-depleted rapid rotator model¹², which assumes membership of a binary system, or for rotationally induced mixing in general¹³. Other explanations might include gravitational settling¹⁴ or diffusion, but these have not operated efficiently in other metal-deficient stars of similar temperature and gravity.

Another surprising result is the high Sr abundance (see Fig. 2). Our observed lower limit for the Sr/Ba ratio ($[\text{Sr}/\text{Ba}]_{\text{non-LTE}} > -0.4$)

suggests that Sr was not produced in the main ‘slow’-neutron-capture (s-) process occurring in asymptotic giant-branch (AGB) stars, given that the typical ratio in Fe-deficient, C-rich and s-process-enhanced stars is¹⁵ $[\text{Sr}/\text{Ba}]_{\text{non-LTE}} < -1.0$. The weak s-process occurring in formerly more massive stars (mass $M > 10M_{\odot}$) might yield Sr with little or no Ba, although theoretical calculations suggest that this process is inefficient at low metallicity¹⁶. Alternatively, the main ‘rapid’-neutron-capture (r-) process might have produced the Sr, given that $[\text{Sr}/\text{Ba}]_{\text{non-LTE}} > -0.4$ is consistent with the ratio observed in metal-poor, r-process-enhanced stars, in which¹⁷ $[\text{Sr}/\text{Ba}]_{\text{non-LTE}} \approx -0.3$ (non-LTE corrections were applied as for HE1327–2326). It is therefore possible that the Sr was formed in a type-II supernova, currently believed to be a site of the main r-process that expelled material from which HE1327–2326 formed. Another possibility is the recently proposed process that might have provided light neutron-capture elements in the early Galaxy¹⁶. Although the site and details of this newly suggested process are not yet known, it might provide an explanation for the excesses of light neutron-capture elements, including Sr, seen in very metal-poor stars. To discriminate the contribution of such processes, further constraints on the Ba abundance, as well as further information on other neutron-capture elements, from higher-quality data, are needed.

With its distinctive chemical composition, HE1327–2326 is an ideal tool for ‘stellar archaeology’ and a better understanding of the early enrichment history, and the first mass function of stars in the Universe. Although two stars, HE1327–2326 and HE0107–5240, are now known with $[\text{Fe}/\text{H}] < -5.0$, none has been found in the range $-5 < [\text{Fe}/\text{H}] < -4$. This is suggestive of an underlying, yet to be

Table 1 Abundance ratios of HE1327–2326

Element	[X/Fe], A(Li), [Fe/H]	
	Subgiant	Dwarf
Li	<1.6	<1.6
C	4.1 ± 0.2	3.9 ± 0.2
N	4.5 ± 0.2	4.2 ± 0.2
O	<4.0	<3.7
Na (LTE)	2.4 ± 0.2	2.4 ± 0.2
Na (non-LTE)	2.0 ± 0.2	2.0 ± 0.2
Mg (LTE)	1.7 ± 0.2	1.7 ± 0.2
Mg (non-LTE)	1.6 ± 0.2	1.6 ± 0.2
Al (LTE)	1.3 ± 0.2	1.3 ± 0.2
Al (non-LTE)	1.7 ± 0.2	1.7 ± 0.2
Ca I	0.1 ± 0.2	0.1 ± 0.2
Ca II	0.9 ± 0.2	0.8 ± 0.2
Ti	0.6 ± 0.2	0.8 ± 0.2
Fe (LTE)	-5.6 ± 0.2	-5.7 ± 0.2
Fe (non-LTE)	-5.4 ± 0.2	-5.5 ± 0.2
Sr (LTE)	1.0 ± 0.2	1.2 ± 0.2
Sr (non-LTE)	1.1 ± 0.2	1.3 ± 0.2
Ba (LTE)	<1.4	<1.7
Ba (non-LTE)	<1.4	<1.7

Using colour– T_{eff} relations²⁴, we determine $T_{\text{eff}} = 6,180 \pm 80$ K from broad-band UBVRI photometry obtained with the MAGNUM telescope²⁵, Hawaii, and JHK magnitudes taken from the 2MASS survey²⁶. To constrain the gravity ($\log g$), we use the proper motion²⁷ ($\mu = 0.073 \pm 0.006$ arcsec yr⁻¹) of HE1327–2326 to set limits on its distance (assuming that the Galactic escape velocity is 500 km s^{-1} and larger than the transverse velocity). It follows that the absolute V magnitude is $M_V > 2.7$ mag which constrains the evolutionary status of the star and therefore $\log g$. Inspection of a 12-billion-year isochrone with $[\text{Fe}/\text{H}] = -3.5$ (the most metal-poor one available²⁸) results in two solutions for the surface gravity $\log g$: 3.7 and 4.5 (in cgs units). Owing to weak observational constraints on $\log g$, it is currently not possible to determine which of these solutions is correct. T_{eff} and $\log g$ are the input for two different plane-parallel LTE model atmospheres that we use to derive the abundances for both gravity solutions: a MARCS model (Gustafsson, B. *et al.*, manuscript in preparation) tailored for the chemical composition of HE1327–2326, and a Kurucz model²⁹. Both sets of abundances agree to within 0.1 dex. Using spectrum synthesis we determine the C and N abundances from CH and NH molecular bands and the upper limit for O from OH molecular bands in the ultraviolet spectral range. Apart from the molecular features, only 13 weak absorption lines are detected in spectral regions not affected by the wings of the Balmer lines and thus suitable for our abundance analysis. Solar abundances were taken from ref. 30. Typical statistical 1σ errors in our abundances are 0.2 dex. Possible systematic errors are judged to be of the same order of magnitude. Non-LTE corrections are included where available²². For the calculation of the non-LTE abundance ratios, the non-LTE iron abundance has been used.

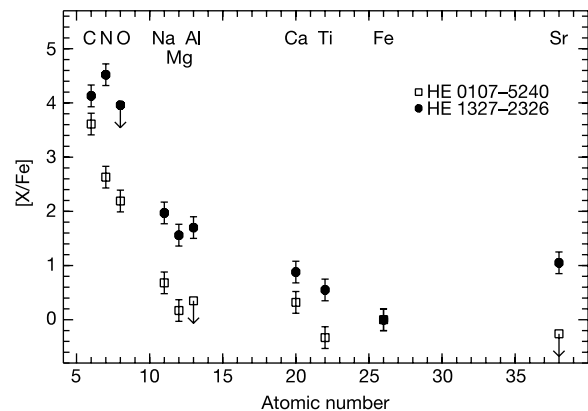


Figure 2 Abundance patterns of HE1327–2326 (subgiant solution, filled circles) and HE0107–5240 (open squares). Typical $1-\sigma$ errors of 0.2 dex are shown in the plot. Upper limits are indicated by an arrow. We adopt the same non-LTE corrections for both stars²², which lead to the modification of the published abundances of HE0107–5240. Consequently, we adopt $[\text{Fe}/\text{H}]_{\text{non-LTE}} = -5.2$ (ref. 9) as the iron abundance for HE0107–5240. These two most Fe-poor stars both have very large C enhancement relative to Fe by a factor of $\sim 5,000$ (HE0107–5240) and $\sim 10,000$ (HE1327–2326). N/Fe is $\sim 30,000$ times the solar value (considering the subgiant solution) in HE1327–2326 whereas it is ~ 200 times the solar value in HE0107–5240. The upper limit for the O abundance of HE1327–2326 is $[\text{O}/\text{Fe}] < 4.0$. Oxygen in HE0107–5240 has recently been determined²³ to be $[\text{O}/\text{Fe}] = 2.3$, which is of the same order as its $[\text{N}/\text{Fe}]$ value. These enormous overabundances in CNO elements suggest that both stars belong to a group of objects sharing a common formation scenario. HE1327–2326 and HE0107–5240 have Ca/Fe and Ti/Fe abundance ratios that are enhanced by factors of less than ten compared to the Sun. The light-element ratios Na/Fe, Mg/Fe and Al/Fe, as well as, surprisingly, Sr/Fe, are all enhanced by factors of ~ 10 to ~ 100 in HE1327–2326. Of these four elements, only Na and Mg are detected in HE0107–5240 with element/Fe ratios close to the solar value. As for several other elements, an upper limit for Ba has been measured in both stars. The Sr/Ba ratio is crucial to identify the origin for the Sr and other heavy elements.

explored, fundamental physical mechanism in the formation of the first stellar objects² which may differ from that for stars more metal-rich than $[\text{Fe}/\text{H}] = -4.0$.

HE1327–2326 permits exploration of the many formation theories of the earliest stars proposed for HE0107–5240. One idea is the pre-enrichment of the gas clouds from which the two stars formed by a first generation ‘faint’ type-II supernova ($M \approx 25M_{\odot}$) experiencing mixing and fallback³, or rotating, massive objects ($M \approx 200\text{--}300M_{\odot}$; ref. 18). In this case, HE1327–2326 would be an early Population II star, formed from gas enriched by one (or possibly a few) of the first type-II supernovae. In particular, the quantitative predictions of abundance patterns by updated models of ‘faint’ type-II supernovae (N. Iwamoto, H. Umeda, N. Tominaga and K. Maeda, personal communication) agree with the chemical abundance patterns of HE1327–2326. The diversity of Mg/Fe ratios found in HE1327–2326 and HE0107–5240 is easily explained by the variety of mixing and fallback. A remaining problem is the Sr, whose production is not yet included in these type-II supernova models.

Another idea is the accretion of heavy elements from the interstellar medium onto first-generation low-mass stars (Population III)^{1,2,19}. In this case, however, the high abundances of lighter elements must be explained by other processes. The unevolved status of HE1327–2326 means that the scenario in which internal mixing processes lead to the dredge-up of processed material can be excluded. A remaining possibility is mass transfer from an AGB star within a binary system^{1,5}. Mass transfer could naturally account for the Li depletion²⁰ found in HE1327–2326. The high Sr abundance is problematic, along with the non-detection of Ba, which cannot be explained by the s-process expected to occur in AGB stars. A crucial test for this scenario is to check the binarity of this object, for which long-period radial velocity monitoring is required. □

Received 22 November 2004; accepted 8 February 2005; doi:10.1038/nature03455.

- Christlieb, N. *et al.* A stellar relic from the early Milky Way. *Nature* **419**, 904–906 (2002).
- Shigeyama, T., Tsujimoto, T. & Yoshii, Y. Excavation of the first stars. *Astrophys. J.* **568**, L57–L60 (2003).
- Umeda, H. & Nomoto, K. First-generation black-hole-forming supernovae and the metal abundance pattern of a very iron-poor star. *Nature* **422**, 871–873 (2003).
- Limongi, M., Chieffi, A. & Bonifacio, P. On the origin of HE 0107–5240, the most iron-deficient star presently known. *Astrophys. J.* **594**, L123–L126 (2003).
- Suda, T., Aikawa, M., Machida, M. N., Fujimoto, M. Y. & Iben, I. Jr. Is HE 0107–5240 a primordial star? The characteristics of extremely metal-poor carbon-rich stars. *Astrophys. J.* **611**, 476–493 (2004).
- Christlieb, N. *et al.* HE 0107–5240, a chemically ancient star. I. A detailed abundance analysis. *Astrophys. J.* **603**, 708–728 (2004).
- Wisotzki, L. *et al.* The Hamburg/ESO survey for bright QSOs. III. A large flux-limited sample of QSOs. *Astron. Astrophys.* **358**, 77–87 (2000).
- Noguchi, K. *et al.* High dispersion spectrograph (HDS) for the Subaru telescope. *Publ. Astron. Soc. Jpn* **54**, 855–864 (2002).
- Beers, T. C. & Christlieb, N. The discovery and analysis of very metal-poor stars in the galaxy. *Annu. Rev. Astron. Astrophys.* (in the press).
- Coc, A., Vangioni-Flam, E., Descouvemont, P., Adahchour, A. & Angulo, C. Updated big bang nucleosynthesis compared with Wilkinson microwave anisotropy probe observations and the abundance of light elements. *Astrophys. J.* **600**, 544–552 (2004).
- Ryan, S. G., Norris, J. E. & Beers, T. C. The Spite lithium plateau: ultrathin but postprimordial. *Astrophys. J.* **523**, 654–677 (1999).
- Ryan, S. G., Gregory, S. G., Kolb, U., Beers, T. C. & Kajino, T. Rapid rotation of ultra-Li-depleted halo stars and their association with blue stragglers. *Astrophys. J.* **571**, 501–511 (2002).
- Pinsonneault, M. H., Walker, T. P., Steigman, G. & Narayanan, V. K. Halo star lithium depletion. *Astrophys. J.* **527**, 180–198 (1999).
- Richard, O., Michaud, G. & Richer, J. Models of metal-poor stars with gravitational settling and radiative accelerations. III. Metallicity dependence. *Astrophys. J.* **580**, 1100–1117 (2002).
- Aoki, W., Norris, J. E., Ryan, S. G., Beers, T. C. & Ando, H. Detection of lead in the carbon-rich, very metal-poor star LP 625–44: A strong constraint on s-process nucleosynthesis at low metallicity. *Astrophys. J.* **536**, L97–L100 (2000).
- Travaglio, C. *et al.* Galactic evolution of Sr, Y, and Zr. A multiplicity of nucleosynthetic processes. *Astrophys. J.* **601**, 864–884 (2004).
- Christlieb, N. *et al.* The Hamburg/ESO R-process enhanced star survey (HERES). I. Project description, and discovery of two stars with strong enhancements of neutron-capture elements. *Astron. Astrophys.* **428**, 1027–1037 (2004).
- Fryer, C. L., Woosley, S. E. & Heger, A. Pair instability supernovae, gravity waves, and gamma-ray transients. *Astrophys. J.* **550**, 372–382 (2001).
- Yoshii, Y. Metal enrichment in the atmospheres of extremely metal-deficient dwarf stars by accretion

of interstellar matter. *Astron. Astrophys.* **97**, 280–290 (1981).

- Norris, J. E., Ryan, S. G., Beers, T. C. & Deliyannis, C. P. Extremely metal-poor stars. III. The Li-depleted main-sequence turnoff dwarfs. *Astrophys. J.* **485**, 370–379 (1997).
- Beers, T. C., Rossi, S., Norris, J. E., Ryan, S. G. & Sheffer, T. Estimation of stellar metal abundance. II. A recalibration of the Ca II K technique, and the autocorrelation function method. *Astron. J.* **117**, 981–1009 (1999).
- Asplund, M. New light on stellar abundances analyses: departures from LTE and homogeneity. *Annu. Rev. Astron. Astrophys.* (in the press).
- Bessell, M. S., Christlieb, N. & Gustafsson, B. On the oxygen abundance of HE 0107–5240. *Astrophys. J.* **612**, L61–L63 (2004).
- Alonso, A., Arribas, S. & Martinez-Roger, C. The empirical scale of temperatures of the low main sequence (F0V–K5V). *Astron. Astrophys.* **313**, 873–890 (1996).
- Yoshii, Y. in *New Trends in Theoretical and Observational Cosmology* (eds Sato, K. & Shiromizu, T.) 235–244 (Universal Academy, Tokyo, 2002).
- Cutri, R. M., *et al.* 2MASS All-Sky Catalog of Point Sources (California Institute of Technology, Pasadena, 2003); (<http://irsa.ipac.caltech.edu/applications/Gator>).
- Girard, T. M. *et al.* The southern proper motion program. III. A near-complete catalog to $V = 17.5$. *Astron. J.* **127**, 3060–3071 (2004).
- Kim, Y., Demarque, P., Yi, S. K. & Alexander, D. R. The Y2 isochrones for alpha-element enhanced mixtures. *Astrophys. J. Suppl.* **143**, 499–511 (2002).
- Kurucz, R. L. *ATLAS9 Stellar Atmosphere Programs and 2 km/s Grid CD-ROM 13* (Smithsonian Astrophysical Observatory, Cambridge, 1993); (<http://kurucz.harvard.edu/cdroms.html>).
- Asplund, M., Grevesse, N. & Sauval, A. J. in *Cosmic Abundances As Records Of Stellar Evolution And Nucleosynthesis* (eds Bash, F. N. & Barnes, T. G.) *ASP Conf. Ser.* (in the press); preprint at (<http://www.arxiv.org/astro-ph/0410214>) (2004).

Acknowledgements We thank A. Steinhauer and C. Thom for obtaining additional observations, N. Iwamoto, K. Maeda, T. Suda, N. Tominaga and H. Umeda for valuable discussions and L. Wisotzki and D. Reimers for leading the HES. This work was supported by the Astronomical Society of Australia (A.F.), Australian Research Council (M.A., A.F., J.E.N.), Ministry of Education, Culture, Sports, Science and Technology in Japan and JSPS (all Japanese co-authors), Deutsche Forschungsgemeinschaft (N.C.), Swedish Research Council (P.S.B., K.E.), US National Science Foundation (T.C.B.) and JINA (T.C.B., N.C., A.F., J.E.N.). This work is based on data collected at the Subaru Telescope, which is operated by the National Astronomical Observatory of Japan.

Competing interests statement The authors declare that they have no competing financial interests.

Correspondence and requests for materials should be addressed to A.F. (anna@mso.anu.edu.au).

Planet–planet scattering in the upsilon Andromedae system

Eric B. Ford¹, Verene Lystad² & Frederic A. Rasio²

¹Department of Astronomy, University of California, Berkeley, California 94720, USA

²Department of Physics and Astronomy, Northwestern University, Evanston 60208, Illinois USA

Doppler spectroscopy has detected 152 planets around nearby stars¹. A major puzzle is why many of their orbits are highly eccentric; all planets in our Solar System are on nearly circular orbits, as is expected if they formed by accretion processes in a protostellar disk. Several mechanisms have been proposed to generate large eccentricities after planet formation, but so far there has been little observational evidence to support any particular model. Here we report that the current orbital configuration of the three giant planets around upsilon Andromedae^{2,3} (υ And) probably results from a close dynamical interaction with another planet⁴, now lost from the system. The planets started on nearly circular orbits, but chaotic evolution caused the outer planet (υ And d) to be perturbed suddenly into a higher-eccentricity orbit. The coupled evolution of the system then causes slow periodic variations in the eccentricity of the middle planet (υ And c). Indeed, we show that υ And c periodically returns to a very nearly circular state every 6,700 years.

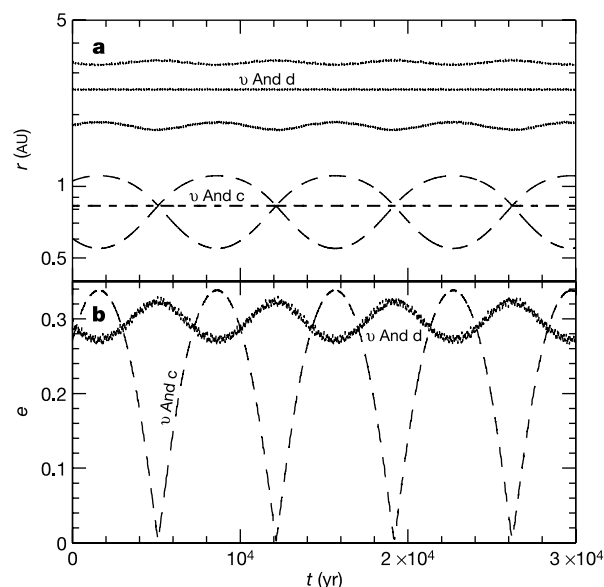


Figure 1 Secular evolution of the planetary system around ν And. **a**, The semimajor axes (middle lines), as well as the periastron and apastron distances (lower and upper lines), for the outer two planets. **b**, The evolution of the orbital eccentricity for each planet. Note that both ν And c (dashed) and ν And d (dotted) have a significant eccentricity at the present time ($t = 0$), but that the eccentricity of ν And c returns periodically to very small values near zero. The results shown here were obtained by direct numerical integration using our best-fit parameters. All direct N -body integrations presented in this paper were performed using Mercury, version 6.1 (ref. 21).

The innermost planet around ν And has a very small orbital eccentricity, as is expected from tidal circularization⁵, but the two outer planets have large eccentricities, both around 0.3. It was quickly recognized after their discovery that the gravitational interaction between these two planets causes significant eccentricity evolution on secular timescales ($\sim 10^4$ years). In particular, in some early solutions, the middle planet's eccentricity appeared to vary periodically with a large amplitude, from a maximum near the present value to a minimum near zero⁶. Later fits to the data suggested that the outer two orbits had arguments of pericentre nearly equal to within about 10° . This could indicate an 'apsidal resonance', in which two elliptical orbits oscillate about an aligned configuration with a small libration amplitude⁷.

Two possible explanations have been proposed for this peculiar orbital configuration. The first is a dynamical mechanism in which a sudden perturbation imparts a finite eccentricity to the outer planet⁴. The subsequent secular evolution causes the middle planet's orbit to become eccentric and can leave the two orbits either circulating, or librating with a large amplitude. In either case, the eccentricity of the middle planet periodically returns to its initial

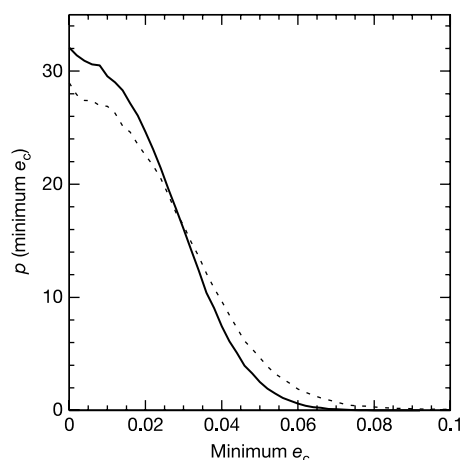


Figure 2 Probability distribution for the minimum eccentricity of ν And c. We draw initial orbital elements for ν And b, c and d from their posterior probability distribution and evolve each system according to classical second-order secular perturbation theory. We find that all allowed orbital solutions have the eccentricity of ν And c oscillating from a maximum value slightly larger than the present value to very nearly zero. The solid curve corresponds to coplanar orbits viewed edge-on. The dotted curve shows the result for orbits with random orientations to the line of sight, but requiring dynamical stability. This implies relative inclinations $< 40^\circ$. Note that systems with relative inclinations $> 140^\circ$ are also dynamically stable. Although such retrograde orbits are unlikely on theoretical grounds, our conclusions permit this possibility. Because the secular perturbation theory averages over the orbits, it is also valid for retrograde orbits. The fact that all allowed solutions result in the eccentricity of ν And c returning to a very small value can be understood easily from lowest-order secular perturbation theory^{4,22}, where the eccentricity vector of each planet can be described as the sum of three rotating eigenvectors in the ($e \cos \omega$, $e \sin \omega$) plane. The eigenvector representing the effects of ν And b on ν And c has a very small amplitude and can be neglected. For the particular configuration of ν And, the two dominant eigenvectors describing ν And c have very nearly the same length. Depending on whether the eigenvector with the faster rotation (higher eigenfrequency) has a slightly larger or smaller length than the other, the vector sum will be rotating around 360° (circulation) or oscillating with an amplitude close to 90° (libration), respectively. Whenever the two vectors are anti-aligned, the magnitude of their sum, that is, the eccentricity of the planet, is very close to zero; when they are aligned, the eccentricity is maximum.

low value. The impulsive perturbation of the outer planet would result naturally from planet–planet scattering, which was one of the earliest mechanisms proposed for inducing large eccentricities in extrasolar planets^{8,9}. In contrast, the second explanation invokes an adiabatic perturbation of the outer planet's eccentricity through torques exerted by an exterior gaseous disk¹⁰. As the eccentricity of the middle planet grows on a similarly long timescale, this would leave the system in an apsidal resonance by damping the libration amplitude to zero. This is a natural extension of more general 'migration scenarios' in which the coupling of a planet's orbit to a gaseous disk could both increase eccentricities¹¹ and lead to orbital decay and the formation of planets with very short orbital periods¹².

The ν And system was the first extrasolar multi-planet system ever discovered by Doppler spectroscopy. Since the first announcement of the outer two planets in 1999, the California and Carnegie Planet Search team has taken over 350 new radial velocity measurements, making ν And one of the systems with the tightest constraints on orbital parameters (Table 1; see also Supplementary Information). The secular evolution of the system is shown in Fig. 1. While the eccentricity of ν And d remains always around 0.3, ν And c returns periodically to a nearly circular orbit with $e < 0.01$. As shown in Fig. 2, this is now a property of all solutions consistent with the data,

Table 1 Masses and orbital parameters for the three planets in ν And

Planet	P (d)	e	ω (deg)	$m \sin i$ (M_J)
b	4.617146(56)	0.016(11)		0.6777(79)
c	241.32(18)	0.258(15)	250.2(4.0)	1.943(35)
d	1301.0(7.0)	0.279(22)	287.9(4.8)	3.943(57)

Results of our new analysis of the ν And radial velocity data^{2,3,20}. We have used the entire Lick Observatory data set, kindly provided to us by D. Fischer. For conciseness, we present only the means and standard deviations on the last two digits (indicated in parentheses) after marginalizing over all other parameters. We list the orbital period (P) in days, the orbital eccentricity (e), the argument of pericentre (ω) in degrees, and the planet mass times the sine of the inclination of the orbital plane to the line of sight ($m \sin i$) in units of Jupiter masses (M_J). More details on our analysis are presented in the Supplementary Information.

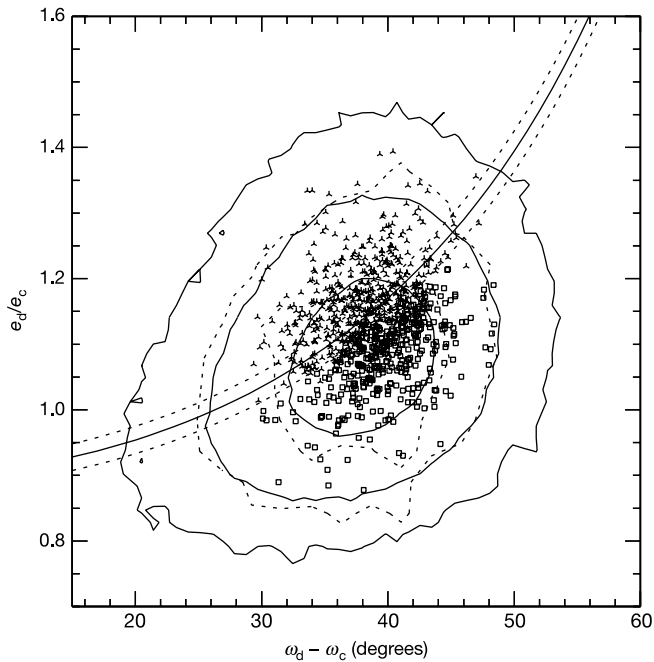


Figure 3 Observational constraints on the secular evolution parameters. The eccentricity ratio is plotted as a function of the difference between the arguments of pericentre for ν And c and d, all at the present epoch. We show the 1-, 2- and 3- σ contours for the posterior probability distribution function marginalized over the remaining fit parameters. The solid contours assume that the radial velocity variations are the result of three non-interacting Keplerian orbits viewed edge-on, while the dotted contours include the mutual gravitational interactions of the planets when fitting to the radial velocity data (only 1- and 2- σ contours are shown). The solid line across shows the separatrix between the librating (upper left) and circulating (lower right) solutions according to the classical second-order perturbation theory (neglecting the inner planet, ν And b). The dotted lines on either side show the variation in the location of the separatrix due to the uncertainty in the remaining orbital elements. The data points show the results of our direct numerical integrations for the full three-planet system: tri-stars (squares) indicate that the system was found to be librating (circulating). Note that, regardless of the assumptions, the separatrix runs right through the 1- σ contours. We find the median libration amplitude of the librating systems to be about 80° .

in contrast to earlier suggestions that it could happen for some solutions^{4,6}.

We have also re-examined the possible presence of apsidal resonance in the system. Remarkably, we find that the allowed solutions all lie very close to the boundary between librating and circulating configurations (Fig. 3). As a consequence, all librating systems have libration amplitudes close to 90° . As shown in Fig. 3, this behaviour is confirmed by direct numerical integrations of all three planets for a large sample of systems covering the allowed parameter space of solutions.

Our results do not change significantly if the assumption of a coplanar system viewed edge-on is relaxed. Inclination angles can be constrained by requiring dynamical stability of the system^{6,7,13,14}. Using the most recent data we find that, for a coplanar configuration of three planets with the best-fit orbital parameters, the system becomes dynamically unstable when $\sin i < 0.5$ (where $\sin i = 1$ corresponds to edge-on). If we relax coplanarity, we find that the system becomes dynamically unstable whenever the relative inclination is greater than about 40° . Throughout the full range of allowed inclination angles we find qualitatively similar behaviour for the secular evolution of the system. For example, Fig. 2 shows that the eccentricity of ν And c would still return periodically to a value near zero. The secular evolution of the system should be re-

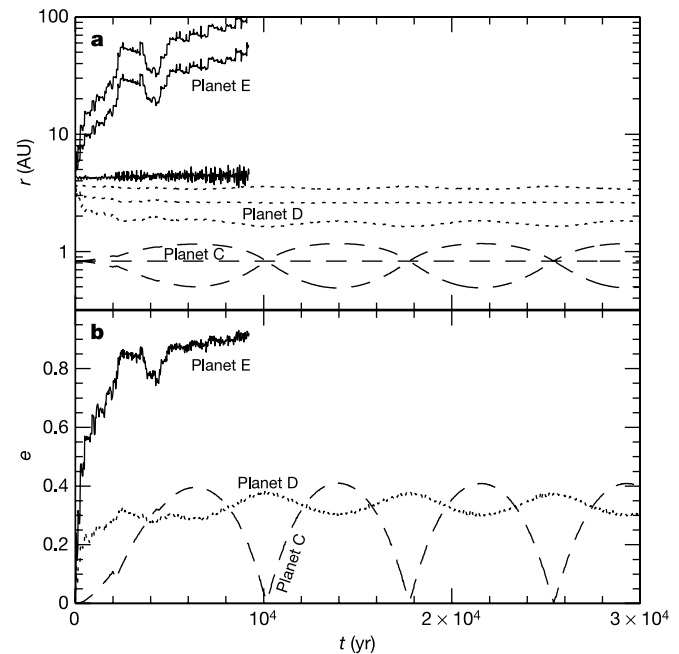


Figure 4 Dynamical evolution of a hypothetical planetary system similar to ν And. Conventions in panels **a** and **b** are as in Fig. 1. Planets B–E are analogous to ν And b, c, d and a hypothetical additional planet. After a brief period of dynamical instability, one planet is ejected, leaving the other two in a configuration very similar to that of ν And c and d. (The innermost planet, B, was not included because it plays a negligible role.) All planets were initially on nearly circular orbits. The initial periods of the two inner planets (C and D) were 241.5 days (equal to the period of ν And c) and 2,100 days. An additional planet (E, solid line), of mass $1.9 M_J$, was placed on an orbit of period 3,333 days, so that the outer two planets were close to their dynamical stability limit²³. The fourth planet (E, solid line) interacted strongly with the third (D, dotted line), while the second planet (C, dashed line) maintained a small eccentricity. After the ejection, the two remaining planets evolve secularly just like ν And c and d (compare with Fig. 1). Although this simulation illustrates a case in which ν And had only one extra planet, our results do not preclude the existence of even more planets at larger distances. In fact, the presence of another planet in an even longer-period orbit could be responsible for triggering the instability after a long period of seemingly ‘stable’ evolution and help raise more quickly the pericentre of the planet that perturbed ν And d. Note that the timescale to completely eject the outer planet from the system (after $\sim 9,000$ years in this particular simulation) is much longer than the timescale of the initial strong scattering. After this initial brief phase of strong interaction, the perturbations on the outer planet are too weak to affect significantly the coupled secular evolution of ν And c and d. Thus, the ‘initial’ eccentricity of ν And c for the secular evolution is determined by its value at the end of the strong interaction phase, rather than that at the time of the final ejection.

evaluated if future observations of ν And were to discover an additional planet (which would be termed ν And e) in a long-period orbit.

Our analysis clearly confirms that the ν And system is evolving exactly as would be expected after an impulsive perturbation to ν And d (ref. 4). The initial sudden change in eccentricity for the outer planet would be naturally produced by a close encounter with another planet, which was ejected from the system as a result. Using our knowledge of planet–planet scattering from several previous studies^{8,15–17}, we determined plausible initial conditions for the original, unstable system. The early dynamical evolution of such a system is illustrated in Fig. 4. After a brief period of strongly chaotic evolution, lasting $\sim 10^3$ years, the outer planet is ejected, and the remaining two planets are left in a dynamical configuration closely resembling that of ν And (see Supplementary Information for a more detailed discussion).

While several other mechanisms (for example, perturbations in a binary star¹⁸, resonances¹⁰, interaction with a gaseous disk¹¹) have been proposed to explain the large eccentricities of extrasolar planets, only planet–planet scattering naturally results in an impulsive perturbation—which is necessary to explain the ν And system. All other mechanisms operate on much longer timescales and would also affect the eccentricity of ν And c, erasing the memory of its initial circular orbit (see Supplementary Information for a more detailed discussion).

Our results have other implications for planet formation. Given the difficulty of forming giant planets at small orbital distances, it is generally assumed that the ν And planets migrated inward to their current locations via interactions with the protoplanetary disk¹². If this is correct, then the small minimum eccentricity of ν And c also provides evidence that its eccentricity at the end of migration had not grown significantly, in contrast to what some theories predict¹¹. However, the possibility of formation *in situ*¹⁹ cannot be excluded by our results. □

Received 18 October 2004; accepted 3 February 2005; doi:10.1038/nature03427.

1. Schneider, J. *Extra-solar Planets Catalog* (<http://cfa-www.harvard.edu/planets/catalog.html>) (2005).
2. Butler, R. P. *et al.* Evidence for multiple companions to Upsilon Andromedae. *Astrophys. J.* **526**, 916–927 (1999).
3. Fischer, D. A. *et al.* A planetary companion to HD 40979 and additional planets orbiting HD 12661 and HD 38529. *Astrophys. J.* **586**, 1394–1408 (2003).
4. Malhotra, R. A dynamical mechanism for establishing apsidal resonance. *Astrophys. J. Lett.* **575**, L33–L36 (2002).
5. Rasio, F. A., Tout, C. A., Lubow, S. H. & Livio, M. Tidal decay of close planetary orbits. *Astrophys. J.* **470**, 1187–1191 (1996).
6. Stepinski, T. F., Malhotra, R. & Black, D. C. The Upsilon Andromedae system: models and stability. *Astrophys. J.* **545**, 1044–1057 (2000).
7. Chiang, E. I., Tabachnik, S. & Tremaine, S. Apsidal alignment in Upsilon Andromedae. *Astron. J.* **122**, 1607–1615 (2001).
8. Rasio, F. A. & Ford, E. B. Dynamical instabilities and the formation of extrasolar planetary systems. *Science* **274**, 954–956 (1996).
9. Weidenschilling, S. J. & Marzari, F. Gravitational scattering as a possible origin for giant planets at small stellar distances. *Nature* **384**, 619–621 (1996).
10. Chiang, E. I. & Murray, N. Eccentricity excitation and apsidal resonance capture in the planetary system Upsilon Andromedae. *Astrophys. J.* **576**, 473–477 (2002).
11. Goldreich, P. & Sari, R. Eccentricity evolution for planets in gaseous disks. *Astrophys. J.* **585**, 1024–1037 (2003).
12. Lin, D. N. C. *et al.* in *Protostars and Planets IV* (eds Mannings, V., Boss, A. P. & Russell, S. S.) 1111–1178 (Univ. Arizona Press, Tucson, 2000).
13. Lissauer, J. J. & Rivera, E. J. Stability analysis of the planetary system orbiting Upsilon Andromedae. II. Simulations using new Lick Observatory fits. *Astrophys. J.* **554**, 1141–1150 (2001).
14. Lystad, V. & Rasio, F. in *The Search for Other Worlds* (eds Holt, S. S. & Deming, D.) 273–276 (AIP Conf. Proc. 713, American Institute of Physics, New York, 2004).
15. Ford, E. B., Havlickova, M. & Rasio, F. A. Dynamical instabilities in extrasolar planetary systems containing two giant planets. *Icarus* **150**, 303–313 (2001).
16. Ford, E. B., Rasio, F. A. & Yu, K. in *Scientific Frontiers in Research on Extrasolar Planets* (eds Deming, D. & Seager, S.) 181–187 (ASP Conf. Ser. 294, Astronomical Society of the Pacific, San Francisco, 2003).
17. Marzari, E. F. & Weidenschilling, S. J. Eccentric extrasolar planets: the jumping Jupiter model. *Icarus* **156**, 570–579 (2002).
18. Holman, M., Touma, T. & Tremaine, S. Chaotic variations in the eccentricity of the planet orbiting 16 Cyg B. *Nature* **386**, 254–256 (1997).
19. Bodenheimer, P., Hubickyj, O. & Lissauer, J. J. Models of the *in situ* formation of detected extrasolar giant planets. *Icarus* **143**, 2–14 (2000).
20. Ford, E. B. Quantifying the uncertainty in the orbits of extrasolar planets. *Astron. J.* (in the press).
21. Chambers, J. E. A hybrid symplectic integrator that permits close encounters between massive bodies. *Mon. Not. R. Astron. Soc.* **304**, 793–799 (1999).
22. Murray, C. D. & Dermott, S. F. *Solar System Dynamics* (Cambridge Univ. Press, New York, 1999).
23. Gladman, B. Dynamics of systems of two close planets. *Icarus* **106**, 247–265 (1993).

Supplementary Information accompanies the paper on www.nature.com/nature.

Acknowledgements We are very grateful to D. Fischer for providing us with the latest radial velocity data on ν And. We also thank E. Chiang, M. H. Lee and S. Peale for discussions. This work was supported by an NSF grant to F.A.R. at Northwestern University and by a Miller Research Fellowship to E.B.F. V.L. acknowledges support from the NASA Undergraduate Summer Research Program at Northwestern. F.A.R. and E.B.F. thank the Kavli Institute for Theoretical Physics for hospitality and support.

Competing interests statement The authors declare that they have no competing financial interests.

Correspondence and requests for materials should be addressed to F.A.R. (rasio@northwestern.edu).

Sensitivity gains in chemosensing by lasing action in organic polymers

Aimée Rose¹, Zhengguo Zhu¹, Conor F. Madigan², Timothy M. Swager¹ & Vladimir Bulović²

¹Department of Chemistry and ²Department of Electrical Engineering and Computer Science, Massachusetts Institute of Technology, Cambridge, Massachusetts 02139, USA

Societal needs for greater security require dramatic improvements in the sensitivity of chemical and biological sensors. To meet this challenge, increasing emphasis in analytical science has been directed towards materials and devices having highly non-linear characteristics; semiconducting organic polymers (SOPs), with their facile excited state (exciton) transport, are prime examples of amplifying materials^{1–3}. SOPs have also been recognized as promising lasing materials⁴, although the susceptibility of these materials to optical damage has thus far limited applications. Here we report that attenuated lasing in optically pumped SOP thin films displays a sensitivity to vapours of explosives more than 30 times higher than is observed from spontaneous emission. Critical to this achievement was the development of a transducing polymer with high thin-film quantum yield, a high optical damage threshold in ambient atmosphere and a record low lasing threshold. Trace vapours of the explosives 2,4,6-trinitrotoluene (TNT) and 2,4-dinitrotoluene (DNT) introduce non-radiative deactivation pathways⁵ that compete with stimulated emission. We demonstrate that the induced cessation of the lasing action, and associated sensitivity enhancement, is most pronounced when films are pumped at intensities near their lasing threshold. The combined gains from amplifying materials and lasing promise to deliver sensors that can detect explosives with unparalleled sensitivity.

One of the successes for SOP-based sensors has been the ultra-trace detection of vapours of TNT and DNT using SOP thin films⁵, which enables the detection of buried landmines on the basis of an explosive vapour signature⁶. The detection of TNT/DNT is facilitated by the electron-deficient-acidic nature of nitroaromatics that bind to the electron-rich semiconducting polymer and quench its fluorescence by an electron transfer mechanism. The emission quenching is enhanced when the excitons rapidly diffuse throughout the SOPs, thereby increasing the probability of an encounter with the TNT/DNT. The low vapour pressure of DNT and TNT (about 100 p.p.b. and 5 p.p.b., respectively) and their efficiency for reversible fluorescence quenching make them ideal analytes to explore novel sensitivity augmentation mechanisms in SOP sensors. Previous approaches to increasing the signal gain in these systems focused on extending the exciton diffusion length by increasing excited state lifetimes and by producing emissive three-dimensional electronic delocalization^{7,8}. Here we demonstrate a different enhancement mechanism that can complement all of these other schemes by harnessing the amplifying nature of the SOP lasing action to generate greater sensitivity. We demonstrate with this new approach a more than 30-fold increase in detection sensitivity when operating near the lasing threshold.

Lasing in both SOPs⁴ and molecular organic materials^{9,10} can generally be obtained for materials with high photoluminescence efficiencies and large Stokes shifts (that minimize the light reabsorption losses)¹¹. However, a limitation of these organic materials has been the lack of durability under the punishing conditions of high pump powers, ambient atmosphere and extended operation, all of which are necessary for many sensory applications. Polymer 1 (Fig. 1) was designed to meet these demanding photophysical and stability needs and also to be sensitive to the chosen analytes, TNT

and DNT. In thin film form, **1** displays a luminescence peak maximum at wavelength $\lambda = 500$ nm. The rapid radiative relaxation (with fluorescence lifetime of ~ 650 ps) contributes to its high fluorescence efficiency in spin-cast thin films of $\Phi = 80\%$, measured by comparison to a standard of 9,10-diphenylanthracene in polymethylmethacrylate ($\Phi = 83\%$). A noteworthy design feature of **1** is the pendant aromatic rings that result in the attachment of hydrocarbon side chains parallel to the polymer backbone, which effectively encapsulate the polymer backbone. We suggest that this encapsulation provides a protective sheath that prevents self-quenching and may provide resistance to photobleaching. The extended π -orbital interactions in **1** create a band structure that facilitates exciton diffusion lengths of 7–15 nm (for studies of exciton diffusion lengths in related poly(phenylene-vinylene)s see ref. 12; see also ref. 13). We observe chemical sensitivity enhancements in the lasing action of **1** when it is fabricated into simple waveguides (Fig. 1a), applied over distributed feedback (DFB) gratings (Fig. 1b), or coated on the exterior of an optical fibre (Fig. 1c). The consistent lasing action sensitivity enhancement for all devices confirms a general principle that is independent of architecture.

Multimode lasing action (see Supplementary Information part D) in thin films of **1** is generated under ambient atmosphere by optically exciting structures with 4-ns nitrogen laser pulses ($\lambda = 337$ nm) at an operating frequency of 30 Hz. The excitation

laser beam was focused into a $9\text{ mm} \times 0.09\text{ mm}$ stripe, and the emission was collected at a 60° angle from the excitation beam, which was incident normal to the substrate. In initial experiments, simple asymmetric waveguides were formed by spin-casting solutions of **1** (50 mg per ml hexane) into thin films with thicknesses ranging from 30 nm to 400 nm on glass substrates. In films thicker than 50 nm, we observe a multimode amplified spontaneous emission (ASE) which peaks at $\lambda = 535$ nm, a wavelength coinciding with the first vibronic transition (0,1) of **1**. The (0,0) transition dominates the spontaneous emission owing to its higher oscillator strength. The selective ASE from the (0,1) mode is probably governed by the reabsorption losses that inhibit ASE from the (0,0) transitions.

Earlier studies demonstrated that TNT and DNT bind tightly to electron-rich SOPs such as **1**, and with short exposures these analytes are localized to the film surface⁴. As polymer **1** exciton diffusion length normal to the surface is 7–15 nm (ref. 14), similar film thicknesses of **1** are needed to ensure that the TNT/DNT exciton quenching at the surface is not overshadowed by unattenuated bulk emission. Such thin SOP films, however, would not sustain the guided light intensities needed for initiation of ASE and lasing action. Planar waveguides were therefore formed by spin-coating a thin film of **1** (with refractive index $n = 1.7$) on a transparent, index-matched thick film of parylene (with $n = 1.67$), which guides most of the photoexcited radiation (Fig. 1a). In these structures, we observed ASE for SOP layers as thin as 40 nm, with thinner films absorbing an insufficient amount of the excitation light. Comparisons of the emission intensities of the (0,0) band at $\lambda = 500$ nm and the (0,1) band at $\lambda = 535$ nm in an 80-nm-thick film of **1** on parylene as a function of pump power reveal the onset of ASE at the integrated threshold power of 80 nW. This corresponds to a threshold peak pump power of $P_{\text{TH}} = 82\text{ W cm}^{-2}$ for each pulse and a pump threshold energy of $E_{\text{TH}} = 330\text{ nJ cm}^{-2}$ per pulse. At these low pumping levels, the signal attenuation due to photobleaching exposure is negligible (see Supplementary Information part C). When a 40-nm-thick film of **1** is deposited onto the DFB structures¹⁵, the threshold energy is further reduced to $E_{\text{TH}} = 40\text{ nJ cm}^{-2}$, corresponding to a peak pump power of $P_{\text{TH}} = 10\text{ W cm}^{-2}$ (Fig. 2), further lowering the associated photo-oxidation.

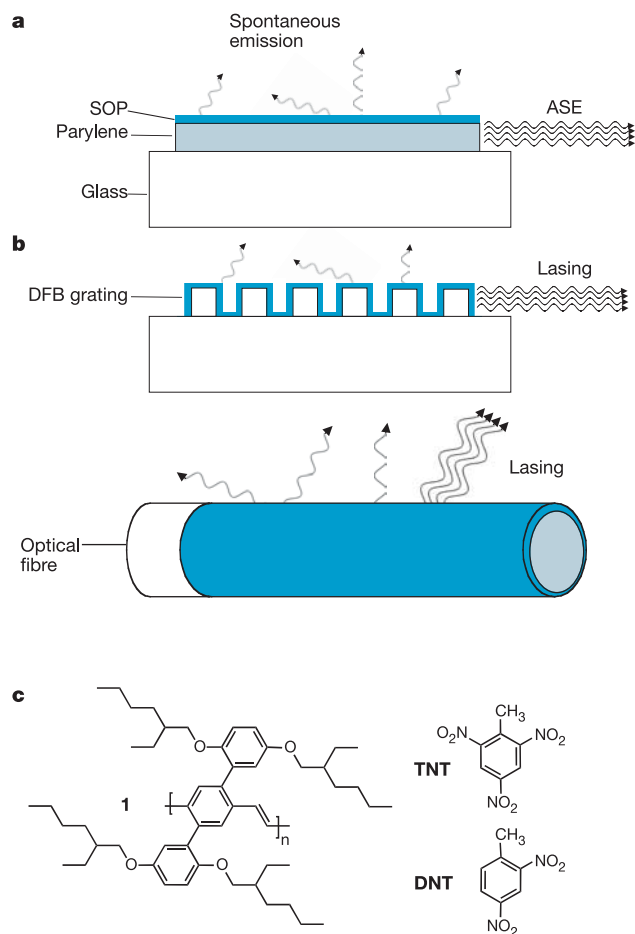


Figure 1 Test structures and materials used to demonstrate lasing chemosensory responses. **a**, Spin-coated polymer **1** films (40–80 nm thick, with refractive index $n = 1.70$) are coated on a transparent 200-nm-thick film of chemical vapour deposited parylene ($n = 1.67$), forming a two-layer index-matched waveguide on glass ($n = 1.45$). **b**, Spin-coated polymer **1** films (40 nm) coated on DFB gratings fabricated from polydimethylsiloxane. **c**, Ring-mode laser structure produced by dip coating polymer **1** on a $25\text{ }\mu\text{m}$ silica optical fibre. Inset, chemical structures of polymer **1**, TNT and DNT.

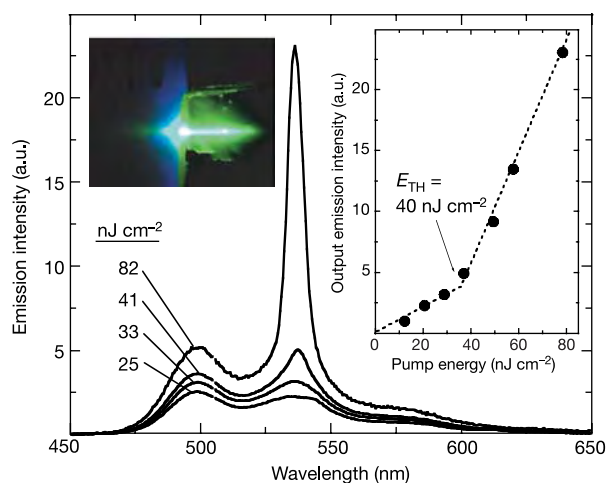


Figure 2 Stimulated and spontaneous emission of polymer **1** atop a polydimethylsiloxane DFB structure, which is shown in Fig. 1b. Main figure, emission spectra of polymer **1** structures at different pump energies show a low stimulated emission threshold (E_{TH}) of 40 nJ cm^{-2} . Right inset, emission intensity of the $\lambda = 535$ nm peak as a function of input power. Left inset, the polymer **1** film is pumped with a strip of $\lambda = 337$ nm pulsed light (right half of the photograph), resulting in a directional emission output (left) above threshold pump energies. a.u., arbitrary units.

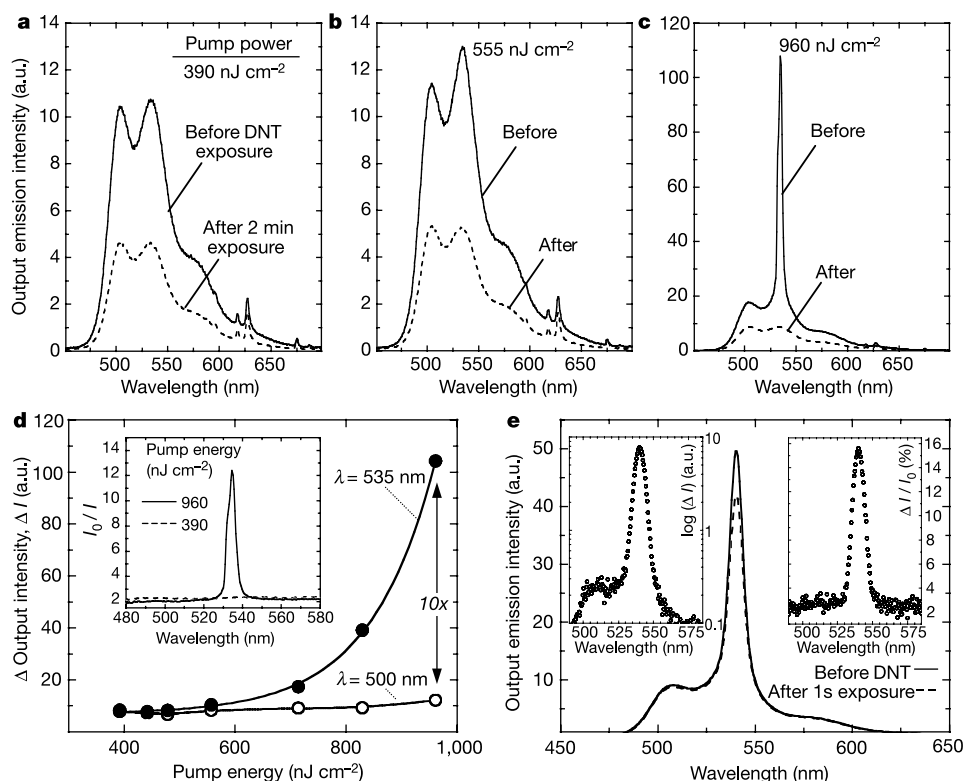


Figure 3 Performance of an optically pumped thin film of polymer **1** on parylene-coated glass before and after exposure to DNT. **a–c**, Spectral response of an optically pumped 80-nm-thick film of polymer **1** on parylene-coated glass (as in Fig. 1a) before and after a 2 min exposure to DNT. The nitrogen pump laser displays pulse-to-pulse power variations, and the approximate pulse energies E (in nJ cm^{-2}) are: **a**, 390; **b**, 560; and **c**, 960. **d**, Change in the output emission intensity, I , at $\lambda = 500$ nm (spontaneous emission; open circles) and $\lambda = 535$ nm (ASE; filled circles) after 2 min DNT exposure as a function of input power. ASE differential caused by quenching shows an increase at higher powers. Inset, emission intensity before exposure to DNT, I_0 , divided by emission

intensity after the exposure, I , plotted for $E = 390$ and 960 nJ cm^{-2} . The largest response occurs at ASE wavelengths ($\lambda = 535$ nm). **e**, Spectral response of a 40-nm-thick film of polymer **1** on parylene-coated glass before and after 1 s exposure to DNT, showing the sizable reduction in the ASE signal with almost no reduction in the spontaneous emission. Plots of the change in the output emission intensity I/I_0 (left inset) and $\Delta I/I_0$ (right inset) show the sensitivity enhancement of 30-fold and fivefold, respectively. The measurement method will therefore determine the ultimate enhancement realized with lasing. The second case represents the minimum enhancement without additional signal processing (that is, averaging).

The ASE performance of polymer **1** compares favourably with previous best demonstrations of stimulated emission in organic materials. The ASE threshold of polymer **1** is significantly lower than the recently reported $E_{\text{TH}} = 3,200 \text{ nJ cm}^{-2}$ (corresponding to a peak pump power of $P_{\text{TH}} = 3.2 \text{ kW cm}^{-2}$) for polyfluorene lasers in a DFB structure¹⁶ and the earlier record of $P_{\text{TH}} = 200 \text{ W cm}^{-2}$ ($E_{\text{TH}} = 400 \text{ nJ cm}^{-2}$) and $P_{\text{TH}} = 85 \text{ W cm}^{-2}$ ($E_{\text{TH}} = 100 \text{ nJ cm}^{-2}$) for doped polymer and doped molecular organic thin films, respectively¹⁷. The improved performance of polymer **1** structures is associated with the improved molecular design that maximizes the thin-film absorption and luminescence efficiency, maximizing the optical gain.

To validate the lasing enhanced sensing principle, we have made comparisons of the sensory responses of the spontaneous emission and ASE of **1** to a static saturated vapour pressure of DNT in air (Fig. 3). The differential sensitivity is readily apparent in a film before and after a 2-min exposure to DNT (Fig. 3a–c), wherein the spontaneous emission at $\lambda = 500$ nm displays approximately a twofold quenching independent of laser power. However, at excitation levels above threshold power, $P_{\text{TH}} = 235 \text{ W cm}^{-2}$ (corresponding to $E_{\text{TH}} = 960 \text{ nJ cm}^{-2}$), the $\lambda = 535$ nm ASE displays a tenfold drop in intensity (Fig. 3d). The differences are even more profound at lower quenching levels afforded by a one-second exposure time of films to saturated DNT vapour, in which we observed a significant attenuation in the ASE peak which is reproducibly >30-fold more sensitive than the attenuation in the spontaneous peak (Fig. 3e).

Whereas signal attenuation due to DNT was readily detected in

these studies, measuring a similar response to TNT vapour proved more challenging. This analyte not only has a lower, easily depleted vapour pressure but also tends to bind tightly to the SOP surface and so is ineffective at quenching excitons generated more than 15 nm deep (the expected exciton diffusion length) in the films. This latter point prompted us to produce ring-mode lasing structures with a thin SOP film by dip-coating a 25- μm -diameter silica fibre with solution of **1** (10 mg per ml hexane). Attenuation in the lasing action on exposure to 5 p.p.b. TNT is observed again while the spontaneous emission of the system remains unchanged (Fig. 4a).

To understand better the origins of the enhanced sensitivity from lasing action, we have modelled the SOP as a standard four-level laser system. The four SOP lasing levels are, in order of increasing energy: (1) the electronic and geometric ground state of the SOP, (2) the electronic ground state and the nuclear excited state, (3) the electronic excited state and nuclear ground state, and (4) the electronic and nuclear excited state of the SOP. From this analysis (see Supplementary Information part B), it is evident that E_{TH} is increased in the presence of TNT and that the chemical sensitivity is determined by the excitation power. The specific considerations are illustrated in a plot of lasing threshold (Fig. 4b). Simple inspection shows that the optimal sensitivity (largest percentage reduction in signal) is obtained when the excitation power of the TNT-quenched polymer is at the onset of ASE. Hence, for ultra-trace detection, the pump power should be only slightly above the E_{TH} of the unquenched polymer.

It is also noteworthy that the difference in emission intensity before and after TNT exposure continues to increase as input power

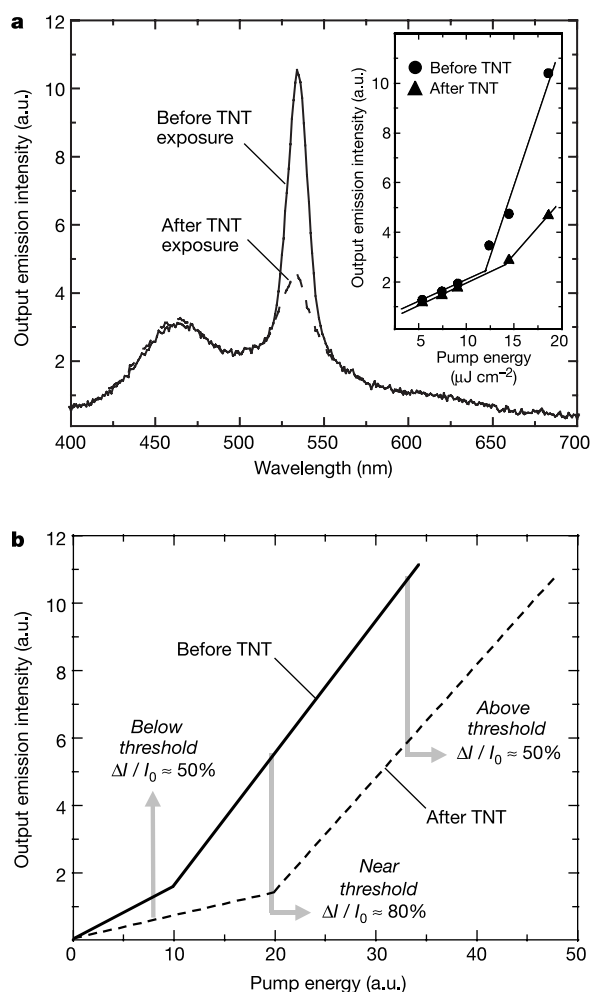


Figure 4 Response of a ring-mode laser coated with polymer **1**. **a**, Spectral response of a ring-mode structure consisting of a 25- μm diameter silica fibre dip-coated with a polymer **1** thin film. ASE attenuation in the absence of spontaneous emission attenuation after 1.5 min exposure to saturated TNT vapour pressure. Inset, plots of ASE peak emission intensity ($\lambda = 535 \text{ nm}$) as a function of excitation power. TNT exposure increases the pump energy threshold. **b**, A general sketch of the ASE E_{TH} increase after exposure of the SOP film to a quencher such as TNT. The change in the output emission intensity divided by the initial output emission intensity, $\Delta I/I_0$, is largest at pump energies just below the E_{TH} of analyte-exposed films.

increases, which results in a quenching signal of larger magnitude. Unfortunately, with longer operating times ($>1 \text{ min}$), the higher excitation power can lead to photobleaching, which interferes with the quenching response. Nevertheless, in instances where the greatest amount of signal is required, one may achieve this through increasing the excitation power, which shortens the operating lifetime of the active polymer film.

The strong binding of the nitroaromatic analytes DNT and TNT to electron-rich SOPs and their ease of reduction makes for a selective response that is relatively immune to interference^{6,18}, and no response was observed upon exposure of our devices to benzene or naphthalene. Current and future efforts focus on incorporating SOPs into high- Q optical feedback structures that will allow further reduction in E_{TH} as well as an enhanced sensitivity with reduced thickness of the active layer. □

Received 28 April 2004; accepted 4 February 2005; doi:10.1038/nature03438.

1. Zhou, Q. & Swager, T. M. Methodology for enhancing the sensitivity of fluorescent chemosensors: energy migration in conjugated polymers. *J. Am. Chem. Soc.* **117**, 7017–7018 (1995).
2. Swager, T. M. The molecular wire approach to sensory signal amplification. *Acc. Chem. Res.* **31**, 201–207 (1998).

3. McQuade, D. T., Pullen, A. E. & Swager, T. M. Conjugated polymer-based sensory materials. *Chem. Rev.* **100**, 2537–2574 (2000).
4. Tessler, N., Denton, G. J. & Friend, R. H. Lasing from conjugated-polymer microcavities. *Nature* **382**, 695–697 (1996).
5. Yang, J.-S. & Swager, T. M. Porous shape persistent fluorescent polymer films: An approach to TNT sensory materials. *J. Am. Chem. Soc.* **120**, 5321–5322 (1998).
6. Cumming, C. J. *et al.* Using novel fluorescent polymers as sensory materials for above-ground sensing of chemical signature compounds emanating from buried landmines. *IEEE Trans. Geosci. Remote Sens.* **39**, 1119–1128 (2001).
7. Rose, A., Lugmair, C. G. & Swager, T. M. Excited-state lifetime modulation in triphenylene-based conjugated polymers. *J. Am. Chem. Soc.* **123**, 11298–11299 (2001).
8. Zahn, S. & Swager, T. M. Three-dimensional electronic delocalization in chiral conjugated polymers. *Angew. Chem. Int. Edn Engl.* **41**, 4225–4230 (2002).
9. Kozlov, V. G., Bulović, V., Burrows, P. E. & Forrest, S. F. Laser action in organic semiconductor waveguide and double heterostructure devices. *Nature* **389**, 362–365 (1997).
10. Bulović, V., Kozlov, V. G., Khalfin, V. B. & Forrest, S. R. Transform-limited, narrow-linewidth lasing action in organic semiconductor microcavities. *Science* **279**, 553–555 (1998).
11. Scherf, U., Riechel, S., Lemmer, U. & Mahr, R. F. Conjugated polymers: lasing and stimulated emission. *Curr. Opin. Solid State Mater. Sci.* **5**, 143–154 (2001).
12. Halls, J. J. M., Pichler, K., Friend, R. H., Moratti, S. C. & Holmes, A. B. Exciton diffusion and dissociation in a poly(*p*-phenylenevinylene)/ C_{60} heterojunction photovoltaic cell. *Appl. Phys. Lett.* **68**, 3120–3122 (1996).
13. Hamer, P. J. *et al.* Optical studies of chemical doping achieved by ion implantation in poly(*p*-phenylene vinylene). *Phil. Mag. B* **73**, 367–382 (1996).
14. Levitsky, I. A., Kim, J. & Swager, T. M. Energy migration in a poly(phenylene ethynylene): determination of interpolymer transport in anisotropic Langmuir-Blodgett films. *J. Am. Chem. Soc.* **121**, 1466–1472 (1999).
15. Kogelnik, H. & Shank, C. V. Stimulated emission in a periodic structure. *Appl. Phys. Lett.* **18**, 152–154 (1971).
16. Heliotis, G. *et al.* Emission characteristics and performance comparison of polyfluorene lasers with one- and two-dimensional distributed feedback. *Adv. Funct. Mater.* **14**, 91–97 (2004).
17. Berggren, M., Dodabalapur, A., Slusher, R. E. & Bao, Z. Light amplification in organic thin films using cascade energy transfer. *Nature* **389**, 466–469 (1997).
18. Yang, Y.-S. & Swager, T. M. Fluorescent porous polymer films as TNT chemosensors: electronic and structural effects. *J. Am. Chem. Soc.* **120**, 11864–11873 (1998).

Supplementary Information accompanies the paper on www.nature.com/nature.

Acknowledgements This work was supported by NASA, the Institute for Soldier Nanotechnologies, and the NSF through the Center for Materials Science and Engineering. A.R. acknowledges V. Sundar and H. Eisler for discussions and J. Ho for providing the DFB structures used in this work.

Competing interests statement The authors declare that they have no competing financial interests.

Correspondence and requests for materials should be addressed to T.M.S. (tswager@mit.edu) or V.B. (bulovic@mit.edu).

Light-induced shape-memory polymers

Andreas Lendlein^{1,2}, Hongyan Jiang^{2,3}, Oliver Jünger^{2,*} & Robert Langer⁴

¹GKSS Research Center Geesthacht GmbH, Institute of Chemistry, Kantstraße 55, D-14513 Teltow, Germany

²Institute for Technology and Development of Medical Devices, RWTH Aachen, Pauwelsstraße 30, and ³mnemoScience GmbH, Pauwelsstraße 19, D-52074 Aachen, Germany

⁴Department of Chemical Engineering, Massachusetts Institute of Technology, 45 Carleton Street, Cambridge, Massachusetts 02139, USA

* Present address: Prof.-Staudinger-Strasse, D-65451 Kelsterbach, Germany

Materials are said to show a shape-memory effect if they can be deformed and fixed into a temporary shape, and recover their original, permanent shape only on exposure to an external stimulus^{1–3}. Shape-memory polymers have received increasing attention because of their scientific and technological significance^{4,5}. In principle, a thermally induced shape-memory effect can be activated by an increase in temperature (also obtained by heating on exposure to an electrical current⁶ or light illumination^{6,7}). Several papers have described light-induced changes in

the shape of polymers^{8–12} and gels^{13–15}, such as contraction^{8–10}, bending^{11–13} or volume changes^{14,15}. Here we report that polymers containing cinnamic groups can be deformed and fixed into pre-determined shapes—such as (but not exclusively) elongated films and tubes, arches or spirals—by ultraviolet light illumination. These new shapes are stable for long time periods, even when heated to 50 °C, and they can recover their original shape at ambient temperatures when exposed to ultraviolet light of a different wavelength. The ability of polymers to form different pre-determined temporary shapes and subsequently recover their original shape at ambient temperatures by remote light activation could lead to a variety of potential medical and other applications.

As an example of the shape-memory functionality of such photoresponsive materials, results of programming and recovering

induced by light are shown in Fig. 1. First, a photoresponsive polymer is conventionally processed into its permanent shape, a plain film (Fig. 1A, a). Afterwards, the polymer film is stretched to 20% elongation by external stress, and both sides of the film were evenly irradiated with ultraviolet (UV) light, wavelength $\lambda > 260$ nm, for 60 min. When the external stress is released, the sample stayed in an elongated form (Fig. 1A, b). Irradiating the elongated shape with UV light of $\lambda < 260$ nm for 60 min induces cleaving of newly formed photosensitive crosslinks. As a consequence, the recovery of the memorized, permanent shape, that is, the original length (Fig. 1A, c), is observed. Other temporary shapes can also be produced. For example, when only the top side of a polymer film in a stretched state (100% elongation) is irradiated with UV light of $\lambda > 260$ nm for 60 min, a corkscrew spiral shape is obtained after release of external stress (Fig. 1B). The formation of such a spiral shape is caused by the formation of two layers. While the deformation is well-fixed for the top layer, the bottom layer keeps its elasticity. As a result, the latter contracts much more than the former when the external stress is released, forming an arch or corkscrew spiral shape.

Thermoresponsive shape-memory polymers consist of two components on the molecular level: first, molecular switches, which are segments having a thermal transition at temperature T_{trans} and that fix the temporary shape by forming physical crosslinks; and second, 'netpoints' that link these segments and determine the permanent shape⁴. In photoresponsive shape-memory polymers (Fig. 1C), we want the netpoints to be part of

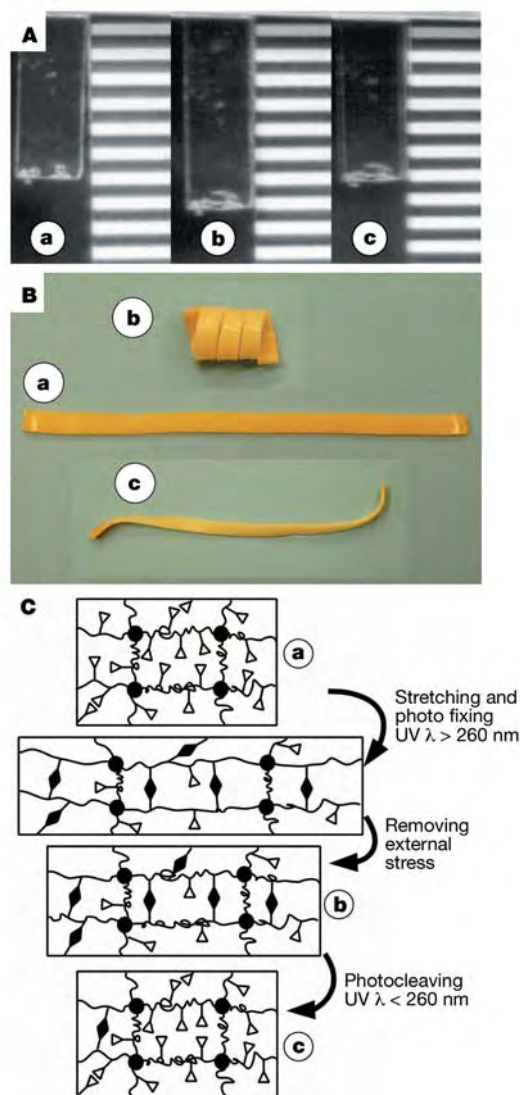


Figure 1 Shape-memory effect of photoresponsive polymers. **A**, A film of grafted polymer BHCA(10,2,1). **a**, Permanent shape (6 cm × 1.2 cm × 0.05 cm); **b**, temporary shape; **c**, recovered permanent shape. **B**, An IPN polymer film. **a**, Permanent shape (8 cm × 0.4 cm × 0.05 cm); **b**, corkscrew spiral temporary shape; **c**, recovered shape obtained by irradiation with UV light of $\lambda < 260$ nm for 60 min. **C**, Molecular mechanism of shape-memory effect of the grafted polymer network: the chromophores (open triangles) are covalently grafted onto the permanent polymer network (filled circles, permanent crosslinks), forming photoreversible crosslinks (filled diamonds); fixation and recovery of the temporary shape are realized by UV light irradiation of suitable wavelengths.

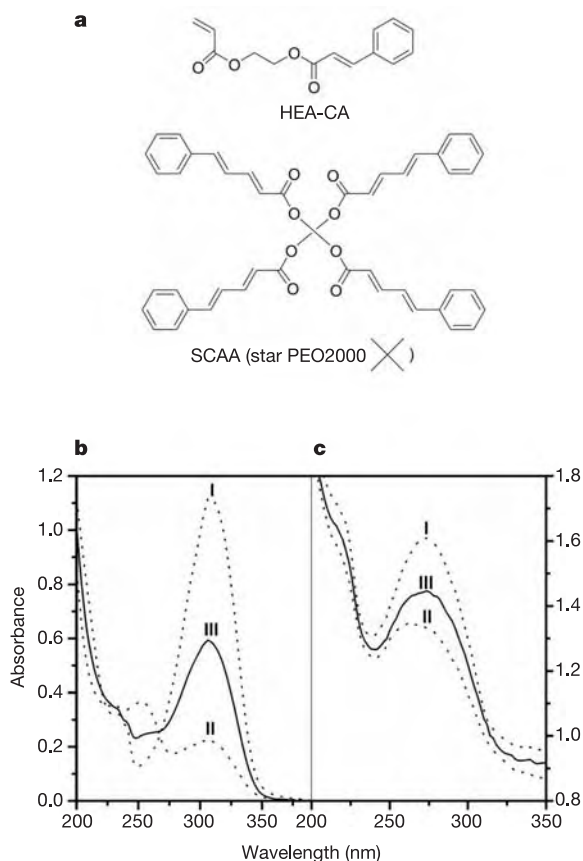


Figure 2 UV absorption spectra. **a**, Chemical structure of molecular switches HEA-CA and SCAA. **b**, A thin film of SCAA. Shown are spectra taken 0 min (spectrum I) and 6 min (II) after exposure to UV light of $\lambda > 260$ nm (crosslinking), and after another 6 min exposure to UV light of $\lambda < 260$ nm (III; de-crosslinking). **c**, A thin film of grafted polymer BHCA(10,2,1): Shown are spectra taken 0 min (I) and 60 min (II) after exposure to UV light of $\lambda > 260$ nm, and after 60 min exposure to UV light of $\lambda < 260$ nm (III).

a covalently crosslinked, amorphous permanent polymer network. The molecular switches are photoresponsive cinnamic acid type molecules such as cinnamic acid (CA)¹⁶ and cinnamylidene acetic acid (CAA)¹⁷, which are able to undergo efficient photoreversible [2 + 2] cycloaddition reactions when exposed to alternating wavelengths (here: $\lambda > 260$ nm or $\lambda < 260$ nm). When such a photo-responsive polymer film (Fig. 1C, a, permanent shape) is stretched, the coiled segments of the amorphous polymer chains between two netpoints are elongated. Upon exposure to UV light of $\lambda > 260$ nm, the elongated segments of the chains are partially fixed owing to formation of new photoresponsive crosslinks, resulting in an elongated new shape (Fig. 1C, b, temporary shape). The photo-responsive crosslinks can be reversibly cleaved by irradiation with UV light of $\lambda < 260$ nm. As a result, the fixed elongated film can recover to the original permanent shape (Fig. 1C, c, recovered permanent shape).

We created two photoresponsive shape-memory polymers. In the first polymer ('grafted polymer'), CA molecules are grafted onto the permanent polymer network, forming a polymer architecture, as schematically illustrated in Fig. 1C. Grafted polymers are obtained by copolymerization of *n*-butylacrylate (BA), hydroxyethyl methacrylate (HEMA) and ethyleneglycol-1-acrylate-2-CA (HEA-CA, in Fig. 2a) with poly(propylene glycol)-dimethacrylate (number-average molecular mass $M_n = 560$ g mol⁻¹, PPG2M560) as crosslinker. Here the molecular switches are photoreversible crosslinks, fixing the deformed temporary shape when irradiated with UV light of $\lambda > 260$ nm, and recovering the shape when exposed to UV light of $\lambda < 260$ nm, as schematically illustrated in Fig. 1C. Only the CA chromophore is selected as a molecular switch, because CAA-containing (meth)acrylates cannot be incorporated into the permanent network via free-radical polymerization. CAA groups may possibly work as free-radical scavengers during the reaction. We synthesized a series of polymer networks with CA groups (Supplementary Table 1). Increasing the amount of HEA-CA increases both the elastic moduli and the glass-transition temperatures (T_g) of the grafted polymer networks, probably owing to the bulky side chains of the CA groups and/or π - π interaction of CA molecules. All polymer networks are amorphous, and have glass-transition temperatures lower than 20 °C.

In the second polymer ('IPN polymer'), the permanent network is formed from BA with 3.0 wt% poly(propylene glycol)-dimethacrylate ($M_n = 1,000$ g mol⁻¹, PPG2M1000) as crosslinker. The permanent network was loaded with about 20 wt% star-poly(ethylene glycol) containing CAA terminal groups (SCAA, in Fig. 2a) by swelling it in a 10 wt% SCAA chloroform solution, forming an opaque yellow film (BP-SCAA) after the solvent is removed. Under the influence of light, these SCAA molecules can be reversibly crosslinked, fixing the temporary shape by interpenetrating the permanent network.

Figure 2b shows changes in the absorption spectra of SCAA when

exposed to UV light of $\lambda > 260$ nm and UV light of $\lambda < 260$ nm. The absorption maximum at 310 nm (Fig. 2b, spectrum I) relating to conjugated diene groups decreased from 1.1 to 0.2 after 6 min exposure to UV light of $\lambda > 260$ nm (Fig. 2b, spectrum II). At the same time, the absorption peak at 250 nm relating to the formed styryl groups increased. The spectral changes are indicative of the disappearance of the double bonds adjacent to carbonyl groups of the CAA moiety, and the formation of a cyclobutane ring bearing styryl groups. On exposure to UV light of $\lambda < 260$ nm for another 6 min, the formed cyclobutane ring is reversibly cleaved and the absorption peak at 310 nm increases again (Fig. 2b, spectrum III). The extent of photoreversibility is around 40%, which is determined from the intensity changes of the absorption peak at 310 nm. The grafted polymer networks having CA groups show a maximum absorption at 275 nm, and a similar photoreversibility (Fig. 2c).

Unlike azo-containing liquid crystal elastomers⁹⁻¹³ but similar to polymers having a thermally induced shape-memory effect²⁻⁴, our photosensitive polymers require pre-determined programming—including deformation into a new shape by application of an external force. Irradiating free-standing, non-deformed polymer films did not change their shape. For potential specific applications, different original and deformed temporary shapes can be made in a pre-determined manner.

The light-induced shape-memory functionality of the polymers was quantified by cyclic photomechanical experiments at 25 or 35 °C under stress-controlled and/or strain-controlled conditions. These cyclic experiments have been designed in analogy to those characterizing a thermally induced effect^{4,18}. A typical elongation–time diagram of the grafted polymer is shown in Fig. 3, which was obtained during a cyclic photomechanical test under stress-controlled conditions. Initially, the sample is elongated to $\epsilon_{\max}$ and the stress is kept constant. During 1.5 h of photo-crosslinking in the first cycle, the elongation stayed constant in the stretched state, indicating no relaxation of the polymer. After switching off the UV light ($\lambda > 260$ nm) and lowering the controlled stress to zero, a temporarily fixed elongation ϵ_u was observed. The fixed temporary shape is maintained for 3 h until UV light of $\lambda < 260$ nm is switched on for 1.5 h, activating the shape recovery of the temporarily fixed shape to nearly its original length characterized by a remaining elongation ϵ_p . This cycle is then repeated twice. The experiment allows the determination of the strain-fixity rate of the *N*th cycle $R_f(N) = \epsilon_u(N)/\epsilon_{\max}$, as well as the strain-recovery rate $R_r(N) = (\epsilon_{\max} - \epsilon_p(N))/\epsilon_{\max}$ ^{18,19}. The strain fixity rate $R_f(N)$ describes the ability to fix the mechanical deformation that has been applied during the photo-crosslinking process. The strain recovery rate $R_r(N)$ quantifies the ability of the material to recover its permanent shape in the *N*th cycle. Similar shape-memory properties were obtained under strain-controlled conditions (Table 1). Increasing the amount of photo-crosslinker HEA-CA improves both R_f and R_r . However, increasing HEA-CA to more

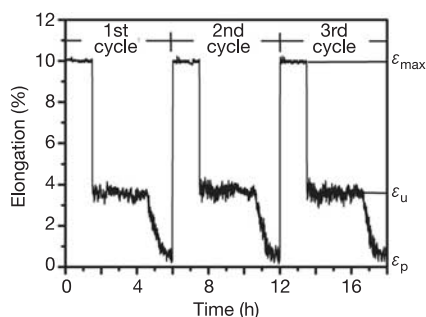


Figure 3 Cyclic photomechanical experiment on a grafted polymer BHCA (10,2,1) under stress-controlled conditions. Here $\epsilon_{\max} = 10\%$, $T = 25$ °C. ϵ_u is the temporarily fixed elongation, and ϵ_p is the remaining elongation after shape recovery.

Table 1 Shape-memory properties of photoresponsive polymer networks at 25 °C

Polymer type	Sample ID	$\epsilon_{\max}$ (%)	Stress control		Strain control	
			R_f (%)	R_r (%)	R_f (%)	R_r (%)
Grafted polymer	BHCA(10,2,0.1)*	10	11	78	14	54§
	BHCA(10,2,0.25)	10	17	82	18	79§
	BHCA(10,2,0.5)	10	25	85	28	82§
	BHCA(10,2,1)	10	36	94	38	97§
	BHCA(10,2,2)	10	36	98	38	98§
	BHCA(10,2,3)	20	n.d.	n.d.	52	95§
IPN polymer	BP-SCAA†	20	20	90	33 (30)‡	98 (95)‡
	BP-SCAA†	50	15	94	20 (20)‡	95 (95)‡
	BP-SCAA†	100	n.d.	n.d.	20 (18)‡	88 (90)‡

Data from third cycle (IPN); fifth cycle (Grafted). n.d., not determined.

*Data in brackets indicate the molar co-monomer ratio (BHCA), with B = BA, H = HEMA, and CA = HEA-CA.

†IPN from BA (B) and 3.0 wt% PPG2M1000 (P) as crosslinker, loaded with 20 wt% SCAA.

‡Data in brackets were obtained at 35 °C.

§ $R_r(N) = (\epsilon_{\max} - \epsilon_p(N))/(\epsilon_{\max} - \epsilon_p(N-1))$

than 10 mol% has no further influence on the shape-memory properties. After two cycles of UV irradiation, the clear film turned pale yellow, possibly owing to photo-Fries rearrangement²⁰, but no changes of shape memory properties were observed (Fig. 3). Although the extent of photoreversibility of the photoresponsive chromophores is around 40%, it is likely that partial cleavage of the photoreversible crosslinks on the microscopic level can lead to the changes of macroscopic mechanical properties and to the shape recovery.

Polymers from the IPN polymer system show an R_f of 20–33% and an R_r of more than 88%. Data obtained at 35 °C (Table 1) are not much different from those obtained at 25 °C. During the irradiation process, the temperature of the polymers does not change, indicating that heating is not the trigger of shape-memory behaviour. When the cyclic photomechanical experiment is performed at lower $\epsilon_{\max}$ (such as 20%), higher values for R_f are obtained. Large deformation will broadly distribute the photo-responsive groups within the specimen under test and make aggregation of these functional groups difficult, thereby influencing the photo-fixing reactions. The relaxation behaviour of IPN polymer networks was investigated in control tests, in which an IPN polymer film was examined in the cyclic mechanical test under the same conditions but without applying UV light. The extent of relaxation determined from these experiments is about 2% in a test with $\epsilon_{\max} = 20\%$, and 5% for $\epsilon_{\max} = 50\%$.

Relative to thermoresponsive shape-memory polymers, the strain-recovery rate of photoresponsive shape-memory polymers is comparable, but the strain-fixity rate is much lower, because the mechanism is different. In photoresponsive polymers, the amorphous permanent network itself has long and coiled segments between two netpoints. During the programming process, the coiled segments of the chains are stretched and elongated (Fig. 1C). The strain-fixation via light irradiation is due to formation of new chemical netpoints, rather than freezing the stretched chains as in the case of thermoresponsive shape-memory polymers. The elastic contraction of the stretched chain segments accounts for the lower strain-fixity rate of light-induced shape-memory polymers. However, the unique characteristics of the photoresponsive shape-memory polymers enable shape-recovery at ambient temperatures by using remote activation, and could eliminate the temperature constraints of thermoresponsive shape-memory polymers for medical and other applications arising from external sample heating. □

Methods

Synthesis

As an example of the grafted polymer, the synthesis of BHCA(10,2,1) (see Table 1 for nomenclature) is described: a solution of 12.80 g (100 mmol) BA, 2.60 g (20 mmol) HEMA, 2.46 g (10 mmol) HEA-CA, 0.22 g (0.4 mmol) PPG2M560 and 0.05 g azo-isobutyronitril (AIBN) was polymerized between two glass sheets with a Teflon spacer of ~0.5 mm at 80 °C for 12 h. The resulting films were swollen in chloroform to remove the residual monomers. SCAA was synthesized by reacting 4-arm star-poly(ethylene glycol) ($M_n = 2,000$, Shearwater Polymers) with an excess of cinnamylidene acetyl chloride in dry THF solution with triethylene amine as the acid receptor.

Cyclic photomechanical and mechanical tests

These were carried out at 25 or 35 °C (selected to be $T > T_g$ and $T > T_{m,SCAA}$, the melting point of SCAA) on a Zwick-Z005, a tensile tester equipped with a 10 N load cell and a thermo-chamber (model 091250, Climatic Systems Ltd) controlled by a Eurotherm 902-904 unit (Eurotherm Controls Ltd). Films are cut into standard samples (ISO 527-2/1BB, length 32 mm, width 2 ± 0.2 mm). The elongation rate is 10 mm min⁻¹. For photomechanical tests, an optical system consisting of two UV lamps (30 W, Cathodeon) and an optical filter having the ability to split a light beam reaching the filter at an angle of 45° were added to this set-up. While light having wavelength $\lambda > 260$ nm is passing through the filter, light with wavelength $\lambda < 260$ nm is being reflected.

The influence of UV irradiation on the temperature of the polymer film in the Zwick thermo-chamber was investigated by inserting a digital thermocouple (RS Components) into the polymer film. During a 1.5 h irradiation period with UV light ($\lambda > 260$ nm or $\lambda < 260$ nm) at 25 or 35 °C, the average fluctuation of the polymer temperature is ± 0.1 °C.

Received 6 December 2004; accepted 2 February 2005; doi:10.1038/nature03496.

- Osada, Y. & Matsuda, A. Shape memory in hydrogels. *Nature* **376**, 219 (1995).
- Liu, C. & Mather, P. T. Thermomechanical characterization of a tailored series of shape memory polymers. *J. Appl. Med. Polym.* **6**, 47–52 (2002).
- Rousseau, I. A. & Mather, P. T. Shape memory effect exhibited by smectic-C liquid crystalline elastomers. *J. Am. Chem. Soc.* **125**, 15300–15301 (2003).
- Lendlein, A. & Kelch, S. Shape-memory polymers. *Angew. Chem. Int. Edn Engl.* **41**, 2034–2057 (2002).
- Wei, Z. G., Sandström, R. & Miyazaki, S. Shape-memory materials and hybrid composites for smart systems: Part I. Shape-memory materials. *J. Mater. Sci.* **33**, 3743–3762 (1998).
- Koerner, H. et al. Remotely actuated polymer nanocomposites—stress-recovery of carbon-nanotube-filled thermoplastic elastomers. *Nature Mater.* **3**, 115–120 (2004).
- Maitland, D. J. et al. Photothermal properties of shape memory polymer micro-actuators for treating stroke. *Lasers Surg. Med.* **30**, 1–11 (2002).
- Eisenbach, C. D. Isomerization of aromatic azo chromophores in poly(ethyl acrylate) networks and photomechanical effect. *Polymer* **21**, 1175–1179 (1980).
- Li, M. H. et al. Light-driven side-on nematic elastomer actuators. *Adv. Mater.* **15**, 569–572 (2003).
- Finkelmann, H., Nishikawa, E., Pereira, G. G. & Warner, M. A new opto-mechanical effect in solids. *Phys. Rev. Lett.* **87**, 015501 (2001).
- Camacho-Lopez, M., Finkelmann, H., Palfy-Muhoray, P. & Shelley, M. Fast liquid-crystal elastomer swims into the dark. *Nature Mater.* **3**, 307–310 (2004).
- Yu, Y. L., Nakano, M. & Ikeda, T. Directed bending of a polymer film by light. *Nature* **425**, 145 (2003).
- Ikeda, T. et al. Anisotropic bending and unbending behavior of azobenzene liquid-crystalline gels by light exposure. *Adv. Mater.* **15**, 201–205 (2003).
- Suzuki, A. & Tanaka, T. Phase transition in polymer gels induced by visible light. *Nature* **346**, 345–347 (1990).
- Juodkazis, S. et al. Reversible phase transitions in polymer gels induced by radiation forces. *Nature* **408**, 178–181 (2000).
- Akabori, S. et al. The novel synthesis of photoreversible cyclobutanocrown ethers by the intramolecular photoaddition of α,ω -dicinnamoyl polyethylene glycol derivatives. *Bull. Chem. Soc. Jpn* **60**, 3453–3455 (1987).
- Tanaka, H. & Honda, K. Photoreversible reactions of polymers containing cinnamylideneacetate derivatives and the model compounds. *J. Polym. Sci. Polym. Chem. Edn* **15**, 2685–2689 (1977).
- Kim, B. K., Lee, S. Y. & Xu, M. Polyurethanes having shape memory effects. *Polymer* **37**, 5781–5793 (1996).
- Tobushi, H., Haray, H., Yamada, E. & Hayashi, S. Thermomechanical properties in a thin film of shape memory polymer of polyurethane series. *Smart Mater. Struct.* **5**, 483–491 (1996).
- Haddleton, D. M., Creed, D., Griffin, A. C., Hoyle, C. E. & Venkataram, K. Photochemical crosslinking of main-chain liquid-crystalline polymers containing cinnamoyl groups. *Makromol. Chem. Rapid Commun.* **10**, 391–396 (1989).

Supplementary Information accompanies the paper on www.nature.com/nature.

Acknowledgements We thank the Bundesministerium für Bildung und Forschung for a BioFuture Award; H.J. is grateful for an AvH fellowship in 1999/2000.

Competing interests statement The authors declare competing financial interests: details accompany the paper on www.nature.com/nature.

Correspondence and requests for materials should be addressed to A.L. (Lendlein@gkss.de).

A doubling of the post-perovskite phase boundary and structure of the Earth's lowermost mantle

John W. Hernlund¹, Christine Thomas³ & Paul J. Tackley^{1,2}

¹Department of Earth and Space Sciences, and ²Institute of Geophysics and Planetary Physics, University of California, Los Angeles, Los Angeles, California 90095-1567, USA

³Department of Earth and Ocean Sciences, University of Liverpool, Liverpool L69 3GP, UK

The thermal structure of the Earth's lowermost mantle—the D'' layer spanning depths of ~2,600–2,900 kilometres¹—is key to understanding the dynamical state and history of our planet. Earth's temperature profile (the geotherm) is mostly constrained by phase transitions, such as freezing at the inner-core boundary or changes in crystal structure within the solid mantle, that are detected as discontinuities in seismic wave speed and for which the pressure and temperature conditions can be constrained by

experiment and theory. A recently discovered phase transition at pressures of the D'' layer^{2–4} is ideally situated to reveal the thermal structure of the lowermost mantle, where no phase transitions were previously known to exist. Here we show that a pair of seismic discontinuities observed in some regions of D'' can be explained by the same phase transition as the result of a double-crossing of the phase boundary by the geotherm at two different depths. This simple model can also explain why a seismic discontinuity is not observed in some other regions, and provides new constraints for the magnitude of temperature variations within D''.

A seismic discontinuity observed in some regions on top of D'' (ref. 5) has previously been ascribed to large thermal gradients caused by subducted oceanic lithosphere, a compositionally distinct layer, or a solid–solid phase transition⁶. Although previous studies lacked knowledge of any phase transition occurring in the mantle at pressures of the D'' region, combined mantle convection and synthetic seismic wave modelling of these potential scenarios suggested that the discontinuity could best be explained by a phase transition having a Clapeyron slope of $\sim 6 \text{ MPa K}^{-1}$ (ref. 7). The recent discovery of a phase transition in MgSiO_3 perovskite (Pv) to a post-perovskite (pPv) phase at pressures of the D'' layer exhibiting a Clapeyron slope of $\sim 7\text{--}10 \text{ MPa K}^{-1}$ (refs 3, 4) is remarkably consistent with these earlier findings.

A thermal boundary layer (TBL) is expected to arise in D'' because heat is transferred across the core–mantle boundary (CMB) via thermal conduction, a relatively inefficient heat-transfer mechanism in silicates. The TBL is characterized by an increased geothermal gradient above the CMB that accounts for potential temperature differences between the ambient mid-mantle and the relatively isothermal outermost core. Unlike other known phase transitions in the Earth's interior, the large Clapeyron slope estimated for the pPv phase boundary combined with the large thermal

gradients expected in a TBL can allow the phase boundary to be crossed by the geotherm at two depths. The pPv phase boundary temperature at CMB pressure (136 GPa) has been estimated to be $\sim 4,000 \text{ K}$ (refs 3, 4), while CMB temperatures estimated from the melting point of iron at the inner-core boundary span a range between $\sim 3,700 \text{ K}$ (ref. 8) and $\sim 4,400 \text{ K}$ (ref. 9). The higher core temperature estimates therefore place the lowermost D'' layer within the Pv stability field, whereas the lower estimates predict that pPv is the dominant phase throughout D'' (Fig. 1a).

Although estimates of the CMB temperature and pPv phase boundary are likely to remain controversial, only the relative differences between these values at the CMB are important in distinguishing the observational consequences of these two scenarios. We call these the single- and double-crossing models, according to the number of times the geotherm intersects the pPv phase boundary. If the pPv phase is to account for the shear-wave velocity (v_s) increase of 2.5–3% that is required to explain the discontinuity on top of D'' (refs 5, 6), then the v_s profiles for a single- or double-crossing will be very different (Fig. 1b). In particular, the double-crossing model predicts the occurrence of two seismic discontinuities of opposite polarity in D'', and an increased overall gradient in seismic velocities in the lowermost TBL. The single-crossing model, on the other hand, predicts a single ubiquitous global discontinuity at the top of D''.

Recently, seismic migration techniques applied to the D'' layer have revealed a pair of oppositely polarized discontinuities beneath Eurasia¹⁰ and the Caribbean region¹¹, in good agreement with the predictions of the double-crossing model. The upper discontinuity coincides with a previously detected discontinuity on top of D'' in both regions^{5,10–11}. Accordingly, we propose a new paradigm for the gross discontinuity structure of D'' in terms of the double-crossing model (Fig. 2a). v_s profiles corresponding to three scenarios illustrate how variations in discontinuity depth relate to variations in mantle temperature (Fig. 2b). Cold mantle temperatures give rise to a thick pPv stability field, whereas a warm mantle causes it to become smaller in extent, and a pPv layer may be entirely absent if mantle temperatures are higher than the pPv transformation temperature at all depths (Fig. 2c). This mechanism suggests a simple explanation for the absence of a seismic discontinuity in some regions⁶.

To test the feasibility of explaining the observed double discontinuity structure with the double-crossing model, we have computed synthetic reflectivity seismograms¹² for the three cases of cold, warm and hot mantle shown in Fig. 2b. In conducting the synthetic tests, we have assumed that the variation of v_s with pressure and temperature are the same in both phases; however, differences are to be expected. Unequal variations should lead to a dependence of discontinuity strength on the depth at which it occurs, but no such correlation is apparent from previous seismic studies spanning a large range of depths ($>300 \text{ km}$) for the upper discontinuity⁶, and therefore this assumption should not significantly affect the results. For comparison, we use data for an event (8 April 1999, at 43.61° N , 130.35° E , depth of 565 km) recorded by the German Regional Seismic Network (GRSN) with corresponding CMB bounce points beneath Eurasia (65° N , 80° E). The data were filtered with a bandpass of 3–50 s to facilitate comparison with the synthetics, which have a dominant period of 5 s.

The synthetic traces for the cold and warm mantle profiles show clear precursors to the core-reflected ScS phase owing to the upper discontinuity (Fig. 3). In the cold mantle case, a weaker intermediate arrival is present only when the lower discontinuity is included in the model, and this phase is also visible in the real data. This intermediate arrival becomes weaker as mantle temperatures are increased, and the pPv layer becomes thinner. To recover the lower discontinuity in both the cold and warm mantle cases, we have migrated both the synthetic and real data using the same methods as in previous studies^{10,11} (Fig. 4). The results of the synthetic tests

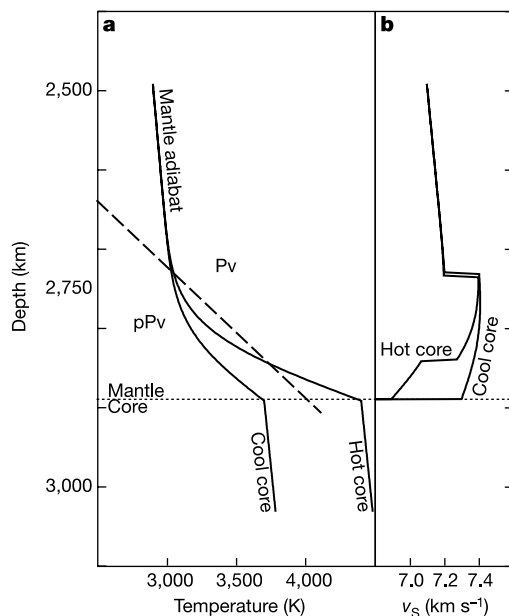


Figure 1 The relationship between the geotherm and pPv phase boundary for different core temperatures (a), and corresponding seismic shear-wave profiles (b). The geotherms (solid lines in a) are plotted following an average mantle adiabat above D'' (ref. 13) with CMB temperatures of 3,700 K (ref. 8; cool core) and 4,400 K (ref. 9; hot core) with the estimated pPv phase boundary (dashed line) superimposed³. The v_s profiles in b are calculated assuming a constant $\partial v_s / \partial T$ of $0.3 \text{ m s}^{-1} \text{ K}^{-1}$ (ref. 18) for both phases, a v_s increase of 0.2 km s^{-1} in the pPv phase at all depths⁶, and an adiabatic variation of v_s referenced to the Preliminary Reference Earth Model¹⁹ and a model geotherm¹³ at a height of 320 km above the CMB.

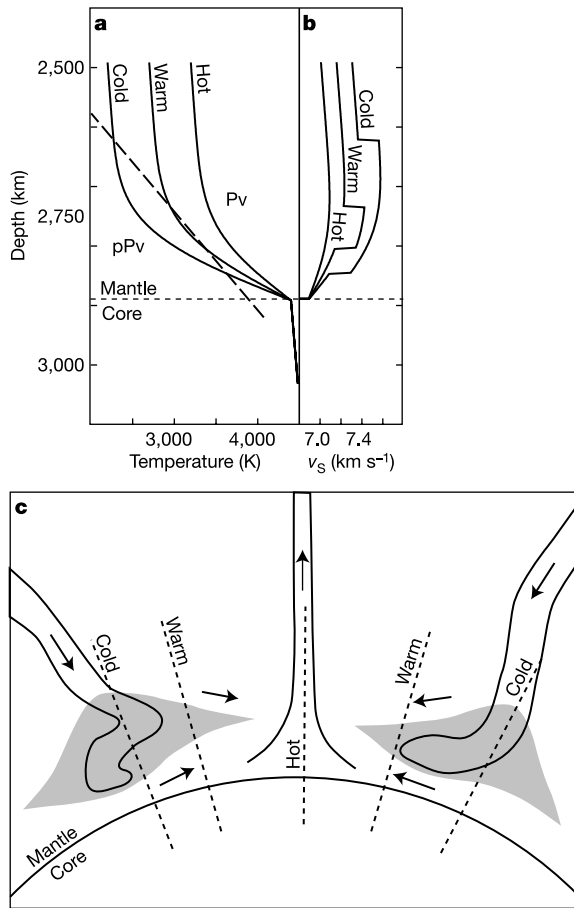


Figure 2 The relationship between three schematic geotherms and the pPv phase boundary (a), corresponding v_s profiles (b), and a sketch of possible lower-mantle structures (c). The v_s profiles are calculated as in Fig. 1b. In c, the pPv layer is shown in light grey, schematic flow directions are indicated by arrows, and several examples of warm, cold and hot mantle profiles are shown (dotted lines).

show that the observed double discontinuity structure is consistent with the double-crossing model. They also indicate that the lower discontinuity will be more difficult to detect in regions where the pPv layer is thin, whereas the upper discontinuity will produce a stronger reflected arrival in all cases. This can explain the lack of a visible lower discontinuity in a portion of the Caribbean region where an upper discontinuity is detected¹¹.

The double-crossing model provides a framework for constraining the magnitude of temperature variations in regions where a D'' discontinuity exists. For example, a Clapeyron slope of 7–10 MPa K⁻¹ translates to a phase boundary gradient of 6–8 K km⁻¹ in D'', which must be a lower bound on the radial temperature gradient in order for a double-crossing to occur (Fig. 2a). A thermal conductivity of 10 W m⁻¹ K⁻¹ (ref. 13) thus yields a minimum conductive heat flux in the range 60–80 mW m⁻², similar to values at the Earth's surface in tectonically active regions. If the thermal gradient is not significantly smaller than this phase boundary gradient even in regions where a double-crossing does not occur, the phase relations predict a global CMB heat flow of the order of 9–13 TW. This rate of cooling is probably more than sufficient to power the dynamo in the outer core that is responsible for producing Earth's magnetic field^{14,15}.

Beneath Eurasia a topographic variation of 55–85 km and 206–316 km above the CMB is observed for the respective lower and upper discontinuities, whereas in the Caribbean region the topographic variations are considerably larger: 66–286 km and 126–416 km above the CMB. The topographic variations imply

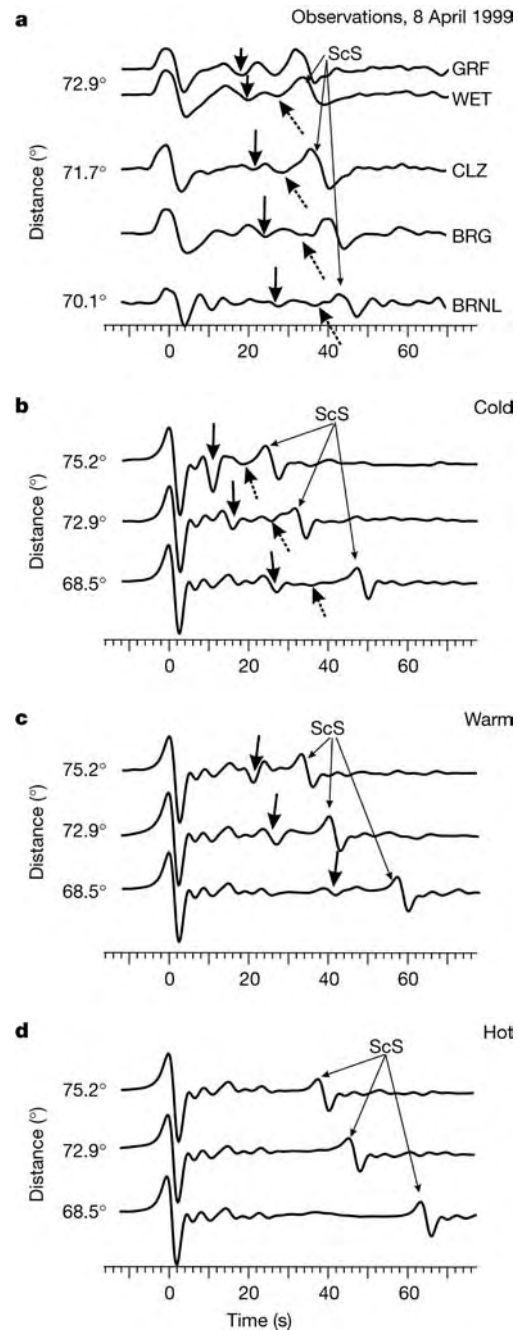


Figure 3 Comparison of data (a) with synthetic shear-wave seismograms calculated for the cold (b), warm (c), and hot (d) v_s profiles shown in Fig. 2b. The model is interpolated to match Earth model ak135²⁰ at depths shallower than 1,000 km, and assumes a density jump of 1% in pPv (refs 2–4). The times are relative, with the traces aligned on the peak of the S phase. The arrival of the core-reflected ScS phase is indicated, the reflection from the upper discontinuity (solid arrows) is shown for the data, cold-mantle and warm-mantle scenarios, and the intermediate arrival attributed to the lower discontinuity (dashed arrows) is shown for the data and cold-mantle scenario.

large lateral temperature gradients along the upper discontinuity of 700–900 K beneath Eurasia and 1,300–1,700 K beneath the Caribbean. These lateral temperature variations are comparable to the expected average temperature change over the TBL, and might be difficult to explain without considering the presence of subducted oceanic lithosphere in both regions. Subducted oceanic lithosphere is expected to exist in both locations if sinking slabs penetrated to the base of Earth's mantle¹⁶. However, other mechanisms for producing the same change in discontinuity depth need to

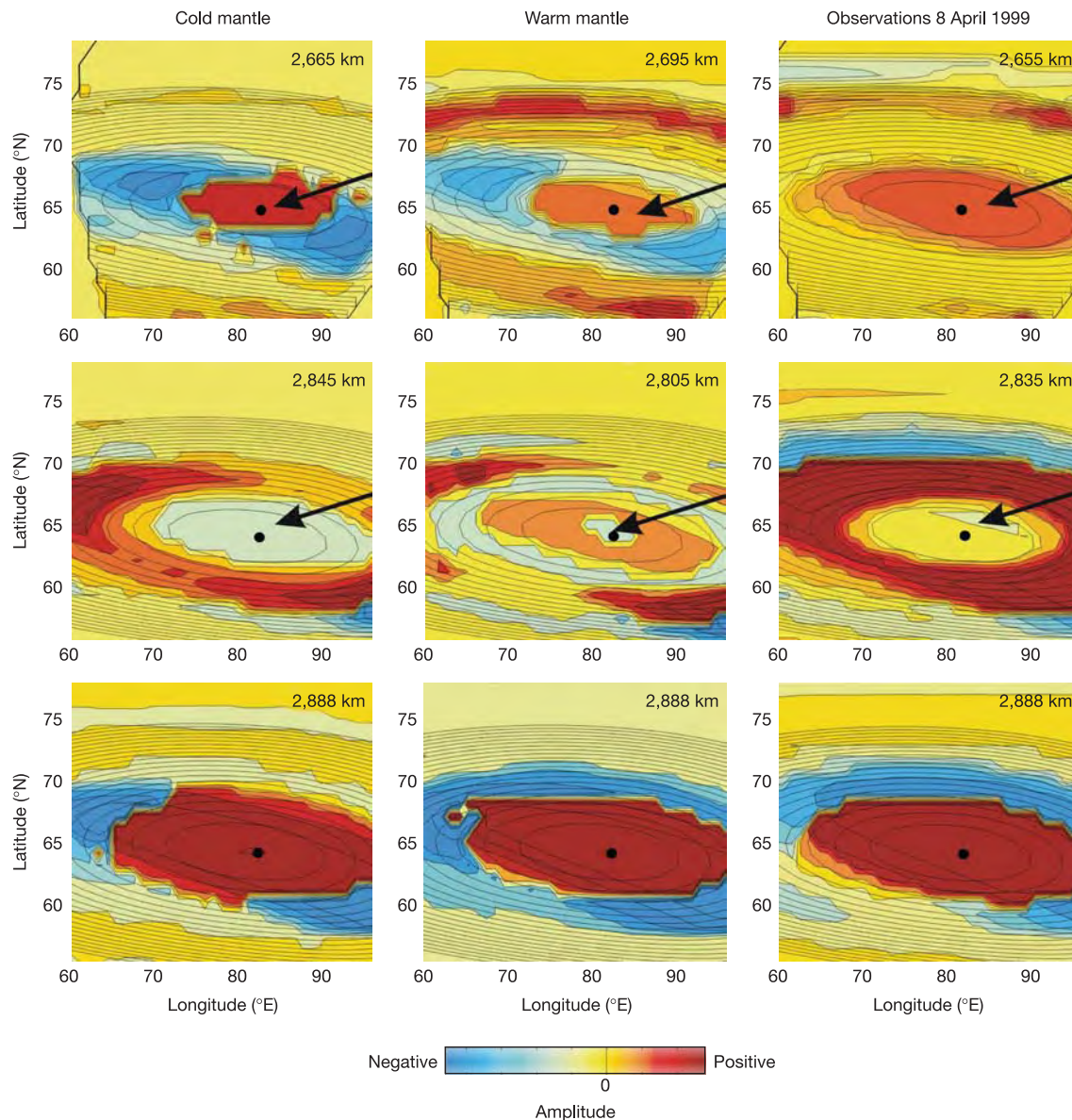


Figure 4 Summary of the migration results for the synthetic cold-mantle (left column) and warm-mantle (centre column) v_s profiles along with real data (right column) using the traces shown in Fig. 3. The discontinuity at the top of D'' is found at two different depths in agreement with the predictions (top row). The deeper discontinuity and its predicted depth

can be recovered in all cases (middle row), but is weaker in the warm mantle. The bottom row shows the migrated core-reflected ScS wave. The colour scale represents the relative amplitude of the migration for each case.

be considered before taking this as conclusive evidence for whole-mantle circulation of subducted lithosphere.

The dynamical consequences of a pPv transition at the top of D'' have already been explored in mantle convection models, where an increased convective instability of the lower TBL has been observed¹⁷. In a double-crossing scenario, the greater vertical extent of the pPv layer in cold regions will increase the relative column density of the mantle in D'' and should provide additional negative buoyancy for cold downwellings such as subducting slabs of oceanic lithosphere. A $\sim 1\%$ density increase associated with the pPv transition^{2–4} might be comparable in magnitude to thermal density changes, and should significantly increase the strength of plate-scale circulation in D'' .

The effects of a two-phase region, due to the presence of iron and aluminium in Pv plus latent heat associated with a dynamical transition, would transform the discontinuity into a gradient, and might change the effective depth/pressure of the phase transition relative to pure MgSiO_3 . Iron is expected to induce a two-phase

region in the phase diagram, in particular because Fe_2O_3 assumes a structure similar to MgSiO_3 -pPv at lower pressures³, but the effects of other chemical components and phases are still largely unknown. A gradient thickness as large as 75 km for the discontinuity on top of D'' is compatible with most seismic studies⁶, so we expect the observable effects of a two-phase region to be contained within a corresponding pressure interval of ~ 4 GPa or less. □

Received 27 September 2004; accepted 15 February 2005; doi:10.1038/nature03472.

1. Bullen, K. E. Compressibility-pressure hypothesis and the Earth's interior. *Mon. Not. R. Astron. Soc. Geophys. Suppl.* **5**, 355–368 (1949).
2. Murakami, M., Hirose, K., Sata, N., Ohishi, Y. & Kawamura, K. Phase transition of MgSiO_3 perovskite in the deep lower mantle. *Science* **304**, 855–858 (2004).
3. Oganov, A. R. & Ono, S. Theoretical and experimental evidence for a post-perovskite phase of MgSiO_3 in Earth's D'' layer. *Nature* **430**, 445–448 (2004).
4. Tsuchiya, T., Tsuchiya, J., Umamoto, K. & Wentzcovitch, R. M. Phase transition in MgSiO_3 perovskite in the earth's lower mantle. *Earth Planet. Sci. Lett.* **224**, 241–248 (2004).
5. Lay, T. & Helmberger, D. V. A shear velocity discontinuity in the lower mantle. *Geophys. Res. Lett.* **10**, 63–66 (1983).

6. Wyssession, M. et al. *The Core-Mantle Boundary Region* 273–298 (American Geophysical Union, Washington, DC, 1998).
7. Sidorin, I., Gurnis, M., Helmberger, D. V. & Ding, X. Interpreting D⁺ seismic structure using synthetic waveforms computed from dynamic models. *Earth Planet. Sci. Lett.* **163**, 31–41 (1998).
8. Boehler, R. High-pressure experiments and the phase diagram of lower mantle and core constituents. *Rev. Geophys.* **38**, 221–245 (2000).
9. Alfe, D., Gillan, M. J. & Price, G. D. Composition and temperature of the Earth's core constrained by combining ab initio calculations and seismic data. *Earth Planet. Sci. Lett.* **195**, 91–98 (2002).
10. Thomas, C., Kendall, J. & Lowman, J. Lower-mantle seismic discontinuities and the thermal morphology of subducted slabs. *Earth Planet. Sci. Lett.* **225**, 105–113 (2004).
11. Thomas, C., Garnero, E. J. & Lay, T. High-resolution imaging of lowermost mantle structure under the Cocos plate. *J. Geophys. Res.* **109**, B08307 (2004).
12. Müller, G. The reflectivity method: A tutorial. *Z. Geophys.* **58**, 153–174 (1985).
13. Stacey, F. *Physics of the Earth* 3rd edn, appendix G (Brookfield, Kenmore, Queensland, 1992).
14. Buffett, B. A. Estimates of heat flow in the deep mantle based on the power requirements for the geodynamo. *Geophys. Res. Lett.* **29**, GL014649 (2002).
15. Christensen, U. R. & Tilgner, A. Power requirement of the geodynamo from ohmic losses in numerical and laboratory dynamos. *Nature* **429**, 169–171 (2004).
16. Lithgow-Bertelloni, C. & Richards, M. A. The dynamics of Cenozoic and Mesozoic plate motions. *Rev. Geophys.* **36**, 27–78 (1998).
17. Nakagawa, T. & Tackley, P. J. Effects of a perovskite-post perovskite phase change near the core-mantle boundary in compressible mantle convection. *Geophys. Res. Lett.* **31**, L16611 (2004).
18. Aizawa, Y. et al. Temperature derivatives of elastic moduli of MgSiO₃ perovskite. *Geophys. Res. Lett.* **31**, L01602 (2004).
19. Dziewonski, A. M. & Anderson, D. L. Preliminary reference earth model. *Phys. Earth Planet. Inter.* **25**, 297–356 (1981).
20. Kennett, B. L. N., Engdahl, E. R. & Buland, R. Constraints on seismic velocities in the Earth from travel times. *Geophys. J. Int.* **122**, 108–124 (1995).

Acknowledgements This collaboration was facilitated by the Meeting of Young Researchers in the Earth Sciences (MYRES) held in La Jolla, California, in August 2004. This work was supported by a grant from IGPP Los Alamos.

Competing interests statement The authors declare that they have no competing financial interests.

Correspondence and requests for materials should be addressed to J.W.H. (hernlund@ess.ucla.edu).

Permian tetrapods from the Sahara show climate-controlled endemism in Pangaea

Christian A. Sidor¹, F. Robin O'Keefe¹, Ross Damiani², J. Sébastien Steyer³, Roger M. H. Smith⁴, Hans C. E. Larsson⁵, Paul C. Sereno⁶, Oumarou Ide⁷ & Abdoulaye Maga⁷

¹Department of Anatomy, New York College of Osteopathic Medicine, Old Westbury, New York 11568, USA

²Bernard Price Institute for Palaeontological Research, University of the Witwatersrand, Johannesburg 2050, South Africa

³Bâtiment de Paléontologie, UMR 5143 CNRS, Département Histoire de la Terre, Muséum national d'Histoire naturelle, 8, rue Buffon, F-75008 Paris, France

⁴South African Museum, Queen Victoria Street, Cape Town 8000, South Africa

⁵Redpath Museum, McGill University, Montreal, Quebec H3A 2K6, Canada

⁶Department of Organismal Biology and Anatomy, University of Chicago, Chicago, Illinois 60637, USA

⁷Institut de Recherches en Sciences Humaines, Niamey, Niger Republic

New fossils from the Upper Permian Moradi Formation of northern Niger^{1–6} provide an insight into the faunas that inhabited low-latitude, xeric environments near the end of the Palaeozoic era (~251 million years ago). We describe here two new temnospondyl amphibians, the cochleosaurid *Nigerpeton ricqlesi* gen. et sp. nov. and the stem edopoid *Saharastega moradiensis* gen. et sp. nov., as relicts of Carboniferous lineages that diverged 40–90 million years earlier^{7–9}. Coupled with a scarcity of therapsids, the new finds suggest that faunas from the poorly sampled xeric belt that straddled the Equator during the Permian period^{10–12} differed markedly from well-sampled

faunas that dominated tropical-to-temperate zones to the north and south^{13–15}. Our results show that long-standing theories of Late Permian faunal homogeneity are probably oversimplified as the result of uneven latitudinal sampling.

For over 150 yr, palaeontologists have understood end-Palaeozoic terrestrial ecosystems largely on the basis of Middle and Late Permian tetrapod faunas from southern Africa. The fauna of these rich beds, particularly South Africa's Karoo Basin, has provided fundamental insights into the origin of modern terrestrial trophic structure¹⁶ and the successive adaptations that set the stage for the subsequent mammalian radiation¹³. A Karoo-like tetrapod fauna has been found across coeval Gondwanan rocks in Brazil, India, Mozambique, Tanzania, Zambia and Zimbabwe. A remarkably similar amniote fauna is known from China, Germany, Laos, Russia and Scotland, although these Laurasian strata host different amphibian groups⁹. The cosmopolitan fauna recorded across Pangaea provides compelling evidence for the unrestricted dispersal of tetrapods during the Middle and Late Permian period. We report here on the discovery of new fossils from West Africa that reveal a highly unusual fauna that has important implications for this interval of vertebrate evolution.

Temnospondyli Zittel, 1888

Edopoidea Romer, 1945

Nigerpeton ricqlesi gen. et sp. nov.

Etymology. *Niger*, for the country, and *herpeton* (Greek), meaning crawler; *ricqlesi*, named for Armand de Ricqlès.

Holotype. MNN MOR69 (Musée National du Niger, Niamey), partial skull and associated atlas vertebra.

Referred material. MNN MOR70, a larger skull preserving most of the left side of the palate, skull roof and lower jaw; MNN MOR83, three isolated sacral neural arches with associated ribs; MNN MOR82, partial femur.

Horizon and locality. Collected from a thin conglomerate in the Upper Permian Moradi Formation, approximately 20 km west of Arlit, north-central Niger.

Diagnosis. Edopoid temnospondyl distinguished from all other edopoids by the unique presence of a highly reduced supratemporal bone, lateral swelling of the maxilla that accommodates two or three fangs medial to the marginal tooth row, maxillary and dentary tooth rows with sporadic appearance of 'doubled' tooth positions, medially positioned premaxillary fangs, and anterior premaxillary vacuities for accommodation of mandibular fangs. Further distinguished from the cochleosaurids *Chenoprosopus* and *Cochleosaurus* by its larger size, extreme preorbital length (~70% of total skull length), anteroposteriorly short skull table, broad sphenethmoid, and the presence of an anterior palatal vacuity.

Saharastega moradiensis gen. et sp. nov.

Etymology. *Sahara*, for the Sahara Desert, and *stega* (Greek), meaning roof; *moradi*, the formation from which the fossil was recovered, and *ensis* (Latin), meaning place or locality.

Holotype. MNN MOR73, nearly complete skull lacking lower jaws.

Horizon and locality. Collected from dark reddish-brown flood-plain deposits of the Moradi Formation, approximately 20 km west of Arlit, north-central Niger.

Diagnosis. Distinguished from all other temnospondyls by the unique presence of an extensive tongue-and-groove articulation between the premaxilla and maxilla, and tabulars with exceptionally large, blunt-ended 'horns' that are directed both laterally and ventrally. Further distinguished from all other edopoids by the following unique combination of characters: orbits broadly separated and close to skull margin, pineal foramen absent, basicranial articulation sutural, fossa subrostralis media present, transverse tooth row present, palatal tusks highly reduced or absent, supraoccipital ossified.

The Upper Permian Moradi Formation of northern Niger is a 100-m-thick succession of fluvial sediments that were deposited as the result of reactivated strike-slip faults bounding the Izégouan-

dane Basin^{1,2,6}. The upper portion of the formation consists of fossiliferous overbank deposits laid down by meandering streams that flowed into a closed, semi-arid continental basin. Although Moradi Formation rivers were sourced in more humid uplands, palaeopedogenic carbonate horizons in the floodplain sediments indicate that basinal conditions were very dry. On a regional scale, palaeoclimatic reconstructions suggest that desert-like conditions prevailed over central and western Africa during Late Permian times^{10–12}. Previously reported members of the Moradi tetrapod fauna include the giant captorhinid *Moradisaurus grandis*^{2,3,5} and the pareiasaur *Bunostegos akokanensis*⁴.

The two amphibians described here are the first temnospondyls known from the Upper Permian of West Africa and the most

primitive temnospondyls from Gondwana. The skull of *Nigerpeton ricqlesi* has crocodilian proportions and size, reaching an adult length of at least 65 cm (Fig. 1a–c, g). In dorsal view (Fig. 1a, b), it is sub-triangular with an elongate, rounded snout and relatively small orbits and external nares. The latter are retracted posteriorly to a position above the maxillary tooth row and open dorsolaterally. The orbits are placed in the posterior one-fifth of the skull and restrict the skull table and its complement of dermal elements to a comparatively small area. As in other cochleosaurids, a characteristic pattern of dermal sculpturing and broad depressions adorns the skull roof. The snout of *Nigerpeton* has paired anterior premaxillary vacuities (av; see Fig. 1) that accommodated the tips of hypertrophied mandibular fangs, a feature common to Triassic

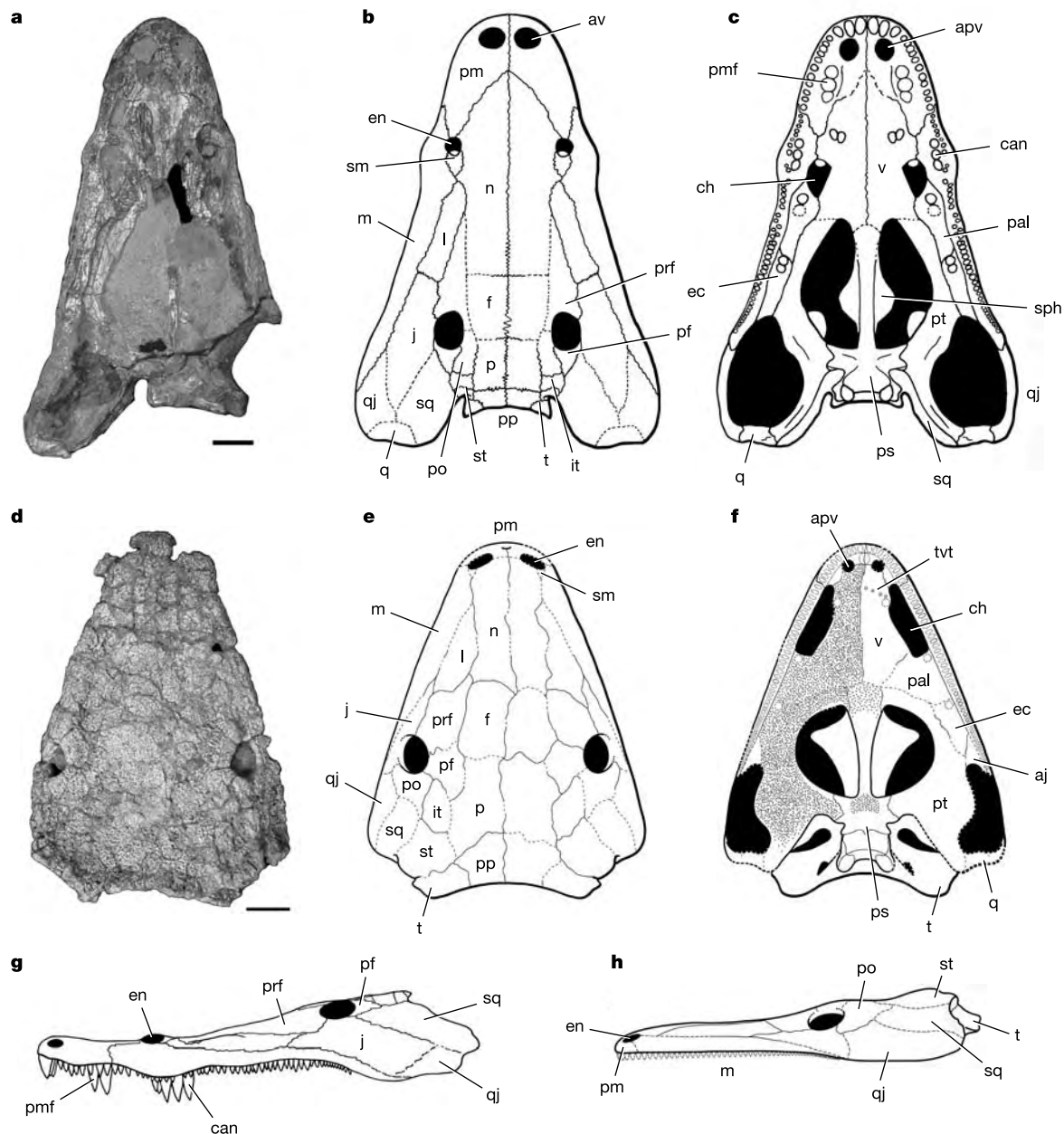


Figure 1 New edopoid temnospondyls from the Moradi Formation. **a**, Photograph of MNM MOR69 in dorsal view. **b, c, g**, Reconstruction of *Nigerpeton ricqlesi* in dorsal (**b**), ventral (**c**) and lateral (**g**) views. **d**, Photograph of MNM MOR73 in dorsal view. **e, f, h**, Reconstruction of *Saharastega moradiensis* in dorsal (**e**), ventral (**f**) and lateral (**h**) views. Scale bars, 5 cm. aj, alary process of jugal; apv, anterior palatal vacuity; av, anterior vacuity in skull roof; can, hypertrophied caniniform region; ch, choana; ec, ectopterygoid;

en, external naris; f, frontal; it, intertemporal; j, jugal; l, lacrimal; m, maxilla; n, nasal; p, parietal; pal, palatine; pf, postfrontal; pm, premaxilla; pmf, premaxillary fang; po, postorbital; pp, postparietal; prf, prefrontal; ps, parasphenoid; pt, pterygoid; q, quadrate; qj, quadratejugal; sm, septomaxilla; sph, sphenethmoid; sq, squamosal; st, supratemporal; t, tabular; tv, transvomerine tooth row; v, vomer.

mastodontosaurids^{17–19} but unknown among primitive temnospondyls, including cochleosaurids. In palatal view (Fig. 1c), *Nigerpeton* displays small, rounded interpterygoid vacuities that are placed in the posterior half of the skull. Anterior to them, the vomers form broad, plate-like elements that separate the posteriorly tapering choanae. The dentition is highly specialized. Anteriorly, the premaxilla bears three enlarged marginal teeth as well as three, medially positioned, palatal fangs. Lateral to the choana, the maxilla has a well-defined swelling that houses three caniniform teeth as well as the smaller teeth of the marginal tooth row. This degree of heterodonty is exceptional among temnospondyls.

The skull of *Saharastega moradiensis* is broadly triangular in outline and shallow in profile, with a moderately long snout (Fig. 1d–f, h). The orbits are posteriorly positioned and face almost wholly laterally. An unusual aspect of the skull is the extreme shortening of the suspensorium, which results in an anteriorly

positioned jaw articulation. As in other edopoids, the pineal foramen and lateral line sulci are absent, the lacrimal is excluded from the anterior margin of the orbit by a contact between the prefrontal and jugal, and an intertemporal ossification is present. The morphology of the tabular of *Saharastega* is markedly different from that of all other temnospondyls; its 'horn' is a very long, robust flange of bone with a blunt end. Orientated ventrolaterally, the horn partially underlaps the posterior margin of the supratemporal. The palate of *Saharastega* has small, rounded interpterygoid vacuities positioned mainly in the posterior half of the skull, paired anterior palatal vacuities, small subtemporal vacuities, and extremely long choanae. A dense field of small denticles uniformly covers the palatal bones, and there is only tenuous evidence for the presence of palatal fangs on the vomers, palatines, or ectopterygoids. *Saharastega* is unique among primitive temnospondyls in possessing a transvomerine tooth row (vtv; see Fig. 1), which is a derived character typical of more advanced temnospondyls^{17–19}.

Several cranial features unite *Nigerpeton* and *Saharastega* with the basal temnospondyl clade Edopoidea^{7–9}. Our phylogenetic analysis indicates that *Saharastega* is the sister taxon of Edopoidea, whereas *Nigerpeton* nests within the clade as the sister taxon to the Late Carboniferous genus *Chenoprosopus* (Fig. 2; see also Supplementary Information for details of analysis). Edopoids were previously restricted to a narrow latitudinal band straddling the palaeoequator in the Late Carboniferous and Early Permian of Euramerica^{7–9}. The discovery of edopoid amphibians in the Moradi Formation implies a stratigraphic gap of at least 40 million years between the early

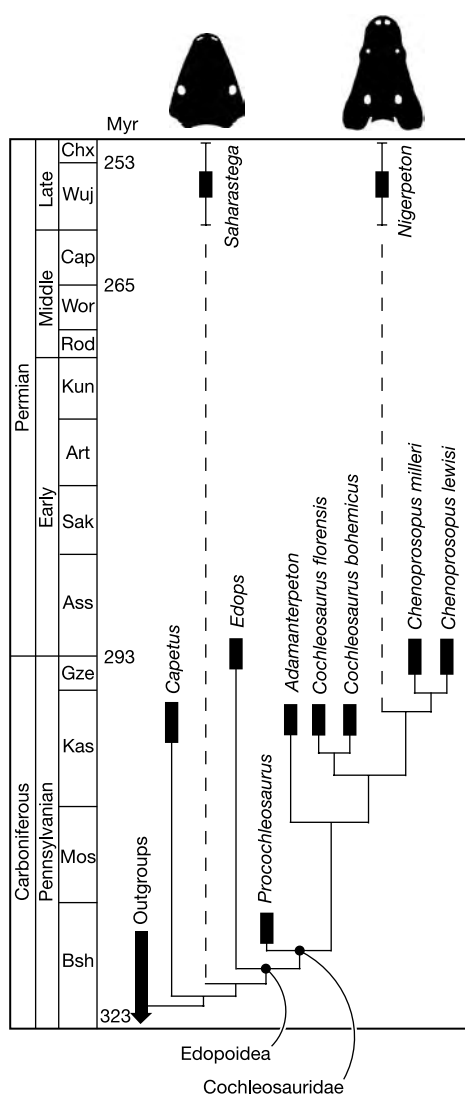


Figure 2 Phylogenetic position of *Nigerpeton ricqlesi* and *Saharastega moradiensis* among basal temnospondyl amphibians based on maximum parsimony analysis (see Supplementary Information for details of analysis). Error bars indicate approximate uncertainty in the age of the Moradi Formation^{2–5}. Stratigraphic distribution of remaining taxa is based on previous studies^{7–9}. Myr, million years. Bsh, Bashkirian; Mos, Moscovian; Kas, Kasimovian; Gze, Gzelian; Ass, Asselian; Sak, Sakmarian; Art, Artinskian; Kun, Kungurian; Rod, Roadian; Wor, Wordian; Cap, Capitanian; Wuj, Wujiapingian; Chx, Changxingian.

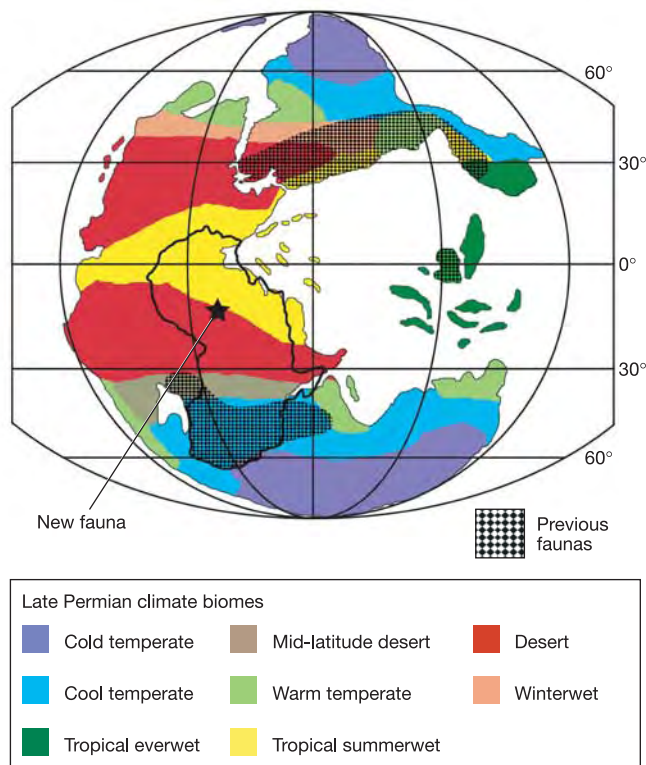


Figure 3 Palaeogeography¹² and palaeoclimatology^{10,11} of the Late Permian showing the geographical distribution of tetrapod faunas^{7–9,13–24}. The distribution of previously recognized, therapsid-dominated faunas is shown by the hatched pattern, and comprises dicynodont herbivores, other therapsid taxa and advanced temnospondyls. The southern range is based on occurrences from South America, southern Africa and India; the northern range is based on European, Russian and Chinese records; the eastern range is based on Laotian records. The new fauna from the Moradi Formation of Niger (star) is characterized by primitive edopoid amphibians, pareiasaurian reptiles and moradisaurine captorhinid reptiles. The modern outline of Africa is highlighted.

appearing Euramerican members of this group and the new African species (Fig. 2). The Moradi amphibians probably represent relicts of a basal group that had long gone extinct elsewhere. In all other regions of Pangaea, more advanced temnospondyls form the bulk of amphibian diversity during the Late Permian, with archegosauroids, dvinosaurians, eryopoids and dissorophoids present in Russia and China, rhinesuchids and archegosauroids present in southern Africa and South America^{9,17–20}, and advanced stereospondyls present in Australia²¹.

Therapsid-dominated faunas composed of the same, or closely related, species have long been known from areas as distant as South Africa and Russia^{13–15,22,23}, and indicate that long-distance, bidirectional north–south exchange was commonplace on Pangaea during the Permian period. In these faunas, dicynodonts are the dominant large-bodied herbivores and advanced temnospondyls represent the principal aquatic predators. Despite its central geographical location (within 15° of the palaeoequator), the fauna from Niger does not overlap at the generic level with the higher-latitude faunas. Dicynodonts have yet to be described from the Moradi Formation^{2–6}, and we found no evidence for their presence, despite intensive fieldwork. Instead, an abundance of captorhinid and pareiasaurian herbivore remains suggests that these taxa were the chief plant consumers^{3–5}.

Climate seems to have had an important role in isolating tetrapods from low latitudes from those at mid- and high latitudes during the Late Permian period (Fig. 3). Geological data and climate simulations suggest that desert-like conditions replaced a more moderate climate in central Pangaea by Middle Permian times^{10–12}. Along with the markedly different floral provinces that characterize northern and southern continents during the Late Permian¹¹, this change in climate may have isolated pockets of a once widespread tetrapod fauna⁹. This hypothesis accounts for the early divergence of the new temnospondyls from Niger, the highly autapomorphic anatomy of the contemporary reptiles^{3–5}, and the similar captorhinid/pareiasaur fauna recovered from low palaeolatitudes in Upper Permian rocks in Morocco^{14,24}.

The degree to which climate influenced the evolutionary biogeography of terrestrial vertebrates towards the end of the Permian period has been previously underestimated because of uneven latitudinal sampling. Mid-to-high-latitude faunas in Russia and South Africa, although fossil-rich and comparatively well known, may have evolved under similar climatic conditions and were probably linked by coastal migration routes. Our discoveries reveal a surprisingly distinctive Permian terrestrial community at low palaeolatitude, and highlight the influence of climate change on large-scale patterns of biotic evolution. □

Received 18 October 2004; accepted 25 January 2005; doi:10.1038/nature03393.

- Ministère des Mines et de l'Hydraulique, Direction des Mines et de la Géologie. *Afaste, Carte Géologique 1:200,000, Notice Explicative* (1977).
- Taquet, P. Un exemple de datation et de corrélation stratigraphique basé sur les Captorhinomorphes (Reptiles cotylosauriens). *Mém. Bureau Recherch. Géol. Min.* **77**, 407–409 (1972).
- de Ricqlès, A. & Taquet, P. La faune de vertébrés du Permien Supérieur du Niger. I. Le captorhinomorphe *Moradisaurus grandis* (Reptilia, Cotylosauria). *Ann. Paléontol.* **68**, 33–106 (1982).
- Sidor, C. A., Blackburn, D. C. & Gado, B. The vertebrate fauna of the Upper Permian of Niger. II. Preliminary description of a new pareiasaur. *Palaeontol. Afr.* **39**, 45–52 (2003).
- O'Keefe, F. R., Sidor, C. A., Larsson, H. C. E., Maga, A. & Ide, O. The vertebrate fauna of the Upper Permian of Niger. III. Ontogeny and morphology of the hindlimb of *Moradisaurus grandis* (Captorhinidae: Moradisaurinae). *J. Vert. Paleontol.* (in the press).
- Taquet, P. *Géologie et Paléontologie du Gisement de Gadoufaoua (Aptien du Niger)* 1–191 (Cahiers de Paléontologie, Paris, 1976).
- Sequeira, S. E. K. The skull of *Cochleosaurus bohemicus*, a temnospondyl from the Czech Republic (Upper Carboniferous) and cochleosaurid interrelationships. *Trans. R. Soc. Edinb. Earth Sci.* **94**, 21–43 (2004).
- Milner, A. R. & Sequeira, S. E. K. A cochleosaurid temnospondyl amphibian from the Middle Pennsylvanian of Linton, Ohio, U.S.A. *Zool. J. Linn. Soc.* **122**, 261–290 (1998).
- Milner, A. R. in *Palaeozoic vertebrate biogeography and biogeography* (ed. Long, J. A.) 324–353 (Belhaven, London, 1993).
- Gibbs, M. T. *et al.* Simulations of Permian climate and comparisons with climate-sensitive sediments. *J. Geol.* **110**, 33–55 (2002).
- Rees, P. M. *et al.* Permian phytogeographic patterns and climate data/model comparisons. *J. Geol.* **110**, 1–31 (2002).

- Ziegler, A. M., Hulver, M. L. & Rowley, D. B. in *Late Glacial and Postglacial Environmental Changes* (ed. Martini, I. P.) 111–146 (Oxford Univ. Press, New York, 1997).
- Rubidge, B. S. & Sidor, C. A. Evolutionary patterns among Permo–Triassic therapsids. *Annu. Rev. Ecol. Syst.* **32**, 449–480 (2001).
- Battail, B. A comparison of Late Permian Gondwanan and Laurasian amniote faunas. *J. Afr. Earth Sci.* **31**, 165–174 (2000).
- Modesto, S. P. & Rybczynski, N. in *The Age of Dinosaurs in Russia and Mongolia* (eds Benton, M. J., Shishkin, M. A., Unwin, D. M. & Kurochkin, E. N.) 17–34 (Cambridge Univ. Press, New York, 2000).
- Bakker, R. T. Anatomical and ecological evidence of endothermy in dinosaurs. *Nature* **238**, 81–85 (1972).
- Damiani, R. J. A systematic revision and phylogenetic analysis of Triassic mastodonsauroids (Temnospondyli: Stereospondyli). *Zool. J. Linn. Soc.* **133**, 379–482 (2001).
- Yates, A. M. & Warren, A. A. The phylogeny of the 'higher' temnospondyls (Vertebrate: Choanata) and its implications for the monophyly and origins of the Stereospondyli. *Zool. J. Linn. Soc.* **128**, 77–121 (2000).
- Schoch, R. R. & Milner, A. R. in *Encyclopedia of Paleoherpetology* (ed. Wellnhofer, P.) Part 3B, 1–203 (Verlag Dr. Friedrich Pfeil, Munich, 2000).
- Dias, E. V. & Barberena, M. C. A temnospondyl amphibian from the Rio do Rasto Formation, Upper Permian of southern Brazil. *An. Acad. Bras. Cienc.* **73**, 135–143 (2001).
- Marsicano, C. A. & Warren, A. A. The first Palaeozoic rhytidosteid record: *Truchoosaurus major* Watson, 1956 from the Late Permian of Australia, and a reassessment of the Rhytidosteidae (Amphibia: Temnospondyli). *Bull. Brit. Mus. Nat. Hist. (Geol.)* **54**, 147–154 (1998).
- Angielczyk, K. D. & Kurkin, A. A. Phylogenetic analysis of Russian Permian dicynodonts (Therapsida: Anomodontia): implications for Permian biostratigraphy and Pangaeian biogeography. *Zool. J. Linn. Soc.* **139**, 157–212 (2003).
- Lucas, S. G. in *Carboniferous and Permian of the World* (eds Hills, L. V., Henderson, C. M. & Bamber, E. W.) 479–491 (Canadian Society of Petroleum Geologists Memoir 19, Calgary, Alberta, 2002).
- Jalil, N.-E. Continental Permian and Triassic vertebrate localities from Algeria and Morocco and their stratigraphical correlations. *J. Afr. Earth Sci.* **29**, 219–226 (1999).

Supplementary Information accompanies the paper on www.nature.com/nature.

Acknowledgements We thank A. Beck, D. Blackburn, J. Conrad, A. Dindine, E. Duneman, B. Gado, T. Lyman, G. Lyon, R. Sadlier and G. Wilson for assistance in the field; A. Crean, E. Love, V. Heisey and J. Groenke for fossil preparation; and S. Spilkevitz for help with Fig. 3. We thank H. Salissou of the Ministère des Enseignements Secondaire et Supérieur, de la Recherche et de la Technologie for permission to conduct fieldwork. We acknowledge the National Geographic Society for support.

Competing interests statement The authors declare that they have no competing financial interests.

Correspondence and requests for materials should be addressed to C.A.S. (casidor@nyit.edu).

Hair cell synaptic ribbons are essential for synchronous auditory signalling

Darina Khimich¹, Régis Nouvian^{1,2}, Rémy Pujol², Susanne tom Dieck^{3*}, Alexander Egner⁴, Eckart D. Gundelfinger³ & Tobias Moser¹

¹Department of Otolaryngology and Center for Molecular Physiology of the Brain, University of Goettingen, 37099 Goettingen, Germany

²INSERM U583–INM, BP 74103 and CRIC, University of Montpellier, 34091 Montpellier Cedex 5, France

³Department of Neurochemistry and Molecular Biology, Leibniz Institute for Neurobiology, 39118 Magdeburg, Germany

⁴Department of NanoBiophotonics, Max Planck Institute for Biophysical Chemistry, Am Fassberg 11, D-37077 Goettingen, Germany

* Present address: Department of Neuroanatomy, Max Planck Institute for Brain Research, 60528 Frankfurt/M, Germany

Hearing relies on faithful synaptic transmission at the ribbon synapse of cochlear inner hair cells (IHCs)^{1–3}. At present, the function of presynaptic ribbons at these synapses is still largely unknown^{1,4}. Here we show that anchoring of IHC ribbons is impaired in mouse mutants for the presynaptic scaffolding protein Bassoon. The lack of active-zone-anchored synaptic ribbons reduced the presynaptic readily releasable vesicle pool, and impaired synchronous auditory signalling as revealed by

recordings of exocytic IHC capacitance changes and sound-evoked activation of spiral ganglion neurons. Both exocytosis of the hair cell releasable vesicle pool and the number of synchronously activated spiral ganglion neurons co-varied with the number of anchored ribbons during development. Interestingly, ribbon-deficient IHCs were still capable of sustained exocytosis with normal Ca^{2+} -dependence. Endocytic membrane retrieval was intact, but an accumulation of tubular and cisternal membrane profiles was observed in ribbon-deficient IHCs. We conclude that ribbon-dependent synchronous release of multiple vesicles at the hair cell afferent synapse is essential for normal hearing.

The afferent synapse of the cochlear IHC is specialized for encoding acoustic signals with high temporal precision over long periods of time¹. Presynaptic active zones in mature IHCs usually contain a single synaptic ribbon—a submicrometre, electron-dense structure tethering synaptic vesicles². Each spiral ganglion neuron (SGN) receives input from only one IHC synapse³. Synaptic ribbons are a hallmark of neurons that transmit graded signals, such as hair cells, retinal photoreceptors and bipolar neurons, pinealocytes and electroreceptors^{1,4,5}. Although some molecular components of the ribbon have recently been identified^{6–8}, the role of synaptic ribbons in transmitter release remains unclear^{1,4,5}. The prevailing view depicts ribbons as conveyor belts for synaptic vesicles, supporting high rates of sustained transmitter release. The investigation of ribbon-deficient synapses, generated by genetic manipulation, offers a direct approach to understanding the function of synaptic ribbons. A mouse mutant for the presynaptic scaffolding protein, Bassoon⁹, lacks synapse-anchored ribbons in retinal photoreceptors and has impaired vision¹⁰. The specific role of the ribbon in transmitter release has remained elusive because synaptic pathophysiology could not be explored at the level of the photoreceptors¹⁰.

Here, we exploit the accessibility of cochlear hair cells for *in vitro* recording¹ to study the consequences of ribbon-deficiency in Bassoon mutants, from the systems response down to the level of the hair cell synapse. Immunohistochemistry of the organs of Corti from 8-week-old mice identified the ribbon-containing IHC synapse as juxtaposed spots of the presynaptic ribbon component RIBEYE⁶ and postsynaptic glutamate receptors (GluR, Figs 1a, g, h and 2g, and Supplementary Movies). Antibodies against the RIBEYE A-domain and against the transcriptional repressor carboxy-terminal binding protein 2 (CtBP2, which is transcribed from the same gene and only differs from the RIBEYE B-domain by 20 amino-terminal amino acids⁶) produced virtually identical staining patterns (Supplementary Fig. 1a). We used the anti-CtBP2 antibody in subsequent experiments because it stained both ribbons and nuclei (Fig. 1a), and therefore allowed us to quantify the number of ribbon-containing synapses per IHC.

A number of Bassoon (Fig. 1a, lower left panel) and Piccolo (Fig. 1a, lower right panel and Supplementary Fig. 1b–d) immunosignals overlapped with IHC ribbons. Some additional, non-overlapping staining of Bassoon and Piccolo probably represented efferent nerve terminals innervating afferent dendrites (axodendritic synapses, Supplementary Fig. 1e, f). Figure 1b is an electron micrograph showing a representative synapse of an 8-week-old wild-type IHC. The ovoid, electron-dense ribbon tethers a halo of synaptic vesicles and is attached to the presynaptic density. A representative IHC synapse from a mutant littermate (Fig. 1c) shows pre- and postsynaptic densities but lacks the ribbon (for quantification of morphological data see Supplementary Table 1). Moreover, the active zone is occupied by tubular and cisternal membrane profiles. These profiles are reminiscent of the large endosomal membrane compartments previously described in ribbon synapses following depolarization^{11–13}, and will be referred to as 'cisterns' (as in ref. 12). A pair of ribbons, embedded in another aggregate of cisterns, was found floating in the cytosol at some

distance from the synapse (Fig. 1c, magnified in Fig. 1d). Numerous cisternal aggregates were observed in IHCs from each of the six mutant mice analysed, whereas only a few, isolated cisterns were found in a fraction (2/7) of wild-type animals. Membrane profiles resembled cisterns when they were distant from floating (Fig. 1d) or anchored ribbons (Fig. 1e), whereas they appeared vesicle-like when close to ribbons. Future experiments will be required to determine the origin and nature of the cisterns in Bassoon-deficient IHCs.

Next, we used the high axial resolution of multifocal 4Pi-microscopy^{14,15} to estimate the ribbon size from large samples of wild-type and mutant RIBEYE immunofluorescent spots. Deconvolved three-dimensional (3D)-reconstructions of RIBEYE- and GluR-labelled afferent synapses showed that both signals were closely juxtaposed in wild-type IHCs (Fig. 1g). In line with the electron microscopy data, mutant IHCs showed predominantly ribbon-deficient synapses (isolated postsynaptic spots in Fig. 1h). As illustrated in the representative 3D-reconstructions of IHC RIBEYE/CtBP2 immunofluorescence (Fig. 1i, j), wild-type IHCs showed many submicrometre RIBEYE spots, whereas mutant IHCs had far fewer, and mainly large, RIBEYE-positive spots

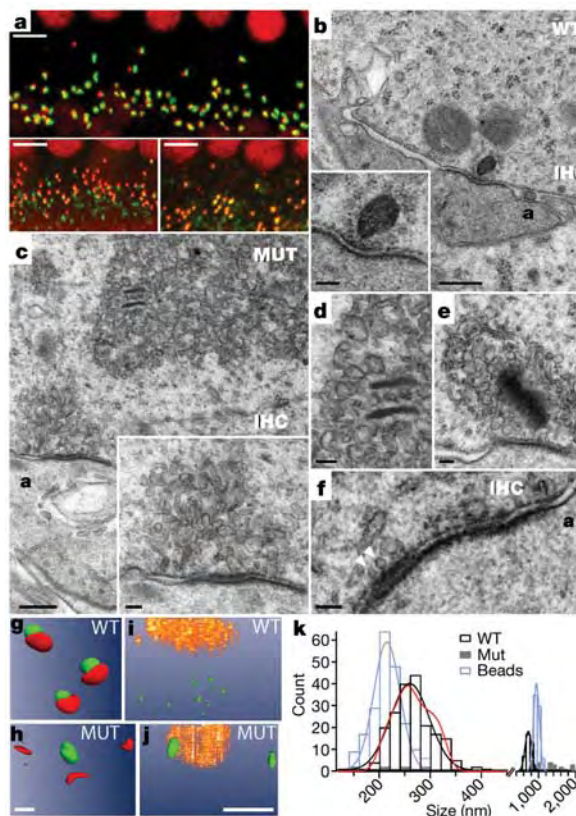


Figure 1 Bassoon anchors synaptic ribbons at IHC active zones. **a**, Confocal stacks of the organ of Corti immunostained for (top panel) RIBEYE (red) and GluR (green), (lower left) RIBEYE and Bassoon (green), or (lower right) RIBEYE and Piccolo (green). **b–f**, Electron micrographs of wild-type (**b**) and mutant (**c–f**) IHC ribbon synapses (shown at higher magnification in insets; the afferent dendrite is marked 'a'). **d**, Floating ribbons in **c** shown at higher magnification. **e**, 'Medusa-like' ribbon of a mutant IHC. **f**, Mutant ribbon-deficient synapse with docked synaptic vesicles (arrowheads). **g–j**, 4Pi microscopy of RIBEYE (green) and juxtaposed postsynaptic GluR (red) spots of wild-type (**g**) and mutant (**h**) afferent synapses after deconvolution. RIBEYE spots (green) of single wild-type (**i**) and mutant (**j**) IHCs (nuclear CtBP2 staining rendered in gold). **k**, Size distributions of 239 wild-type (black) and 43 mutant (grey) RIBEYE spots, as well as of latex beads (blue) with 216- and 1,000-nm diameters. Red line shows ellipsoid model, other lines show spherical model (gaussian fits). Scale bars, 5 μm (**a**, **i** and **j**) 500 nm (**g**, **h**), 400 nm (**b**, **c**), 100 nm (insets to **b** and **c**; **d–f**).

(65% large spots compared with 16% in wild type). The estimates of ribbon size obtained from mutant and wild-type RIBEYE spots (described in Supplementary Methods) are plotted in Fig. 1k. The size distribution of wild-type ribbons had a peak at 260 nm. It was approximated by a model that assumed an ellipsoid shape of the average wild-type ribbon and a random orientation of the ribbon with respect to the optical axis. The three principal axes of the average wild-type ribbon were estimated as 201, 255 and 332 nm (red line in Fig. 1k, Supplementary Table 1 and Supplementary Methods). The large RIBEYE spots in mutant and wild-type IHCs probably corresponded to stacks of floating ribbons, as observed using electron microscopy (for example, Fig. 1d).

Next, we measured cochlear responses in 8-week-old mice to quantify the consequences of IHC ribbon deficiency for auditory function. Figure 2 documents the analysis of a representative pair of wild-type and mutant littermates. First, we recorded auditory brainstem responses to clicks and short tone bursts in order to assess synchronized neural activity in the ascending auditory pathway (Fig. 2a, b). Responses from the mutant showed distortion of auditory brainstem response waveforms and increased thresholds. The first peak (I) was reduced and delayed for supra-threshold stimuli, reflecting impaired compound action potential in the mutant spiral ganglion. To ensure that this reduction of synchronous SGN activation was not due to defects in outer hair cell amplification, we also measured distortion product otoacoustic emissions¹⁶, and found them to be similar in both mutant and wild type (Fig. 2c, d). Thus, IHC ribbon-deficiency caused reduced neural output despite intact outer hair cell function, as confirmed by electrocochleography and otoacoustic emissions in additional experiments (Supplementary Fig. 2). A combination of pathological auditory brainstem responses with intact otoacoustic emissions defines human auditory neuropathy, a peripheral auditory disorder that is associated with poor speech discrimination. The cellular mechanisms of auditory neuropathy are not yet understood, but may involve defects of IHCs or SGNs¹⁷.

Next, we performed a patch-clamp analysis of presynaptic IHC function, using the same mice to test the hypothesis that synaptic dysfunction underlies their auditory neuropathy. Fast exocytosis (indicated by membrane capacitance increments in response to short stimuli¹⁸) was strongly decreased in mutant IHCs (Fig. 2e, f), and Ca^{2+} current was reduced. The organs of Corti were immunostained in order to quantify the number of ribbon-containing synapses (Fig. 2g), as introduced in Fig. 1a and illustrated by Supplementary Movies 1a and b. A strong reduction in synapse-anchored ribbons was observed in the mutant. Both genotypes showed comparable numbers of postsynaptic spots. The morphological findings are summarized in Supplementary Table 1, and Fig. 2h correlates the morphological and physiological results obtained from five mutant and seven wild-type mice. Data from two wild-type and three mutant mice were excluded because of uncertain outer hair cell function (poor otoacoustic emissions). In summary, the lack of synapse-anchored ribbons caused a reduction in fast exocytosis and in synchronous synaptic activation of SGNs.

This conclusion was further supported by three observations. First, we observed larger numbers of synapse-anchored ribbons in both wild-type and mutant IHCs of 3-week-old mice (just after the onset of hearing), which correlated with increased presynaptic IHC and postsynaptic SGN responses when compared with 8-week-old mice of the same genotype. The reduction in anchored ribbons probably reflected developmental changes in the number of IHC ribbons¹⁹. Figure 3a correlates fast IHC exocytosis (membrane capacitance increments in response to 10-ms depolarization) and SGN compound action potential amplitudes with the number of synapse-anchored ribbons over all four groups. Ca^{2+} currents were significantly reduced in mutant IHCs of either age (Fig. 3a), possibly owing to insufficient recruitment and/or stabilization of $\text{Ca}_v1.3$ channels at the active zone. They showed, however, a normal

voltage-dependence of activation (Supplementary Fig. 3a). This impaired Ca^{2+} influx might contribute to the synaptic defect. However, Ca^{2+} currents were comparable between 3- and 8-week-old IHCs of each genotype, yet their synaptic functions differed. This argues for a genuine dependence of hair cell synaptic function on the presence of functional ribbons. Furthermore, we did not observe the mutant exocytic phenotype when we reduced the Ca^{2+} current in 8-week-old wild-type IHCs to the level in the mutants by

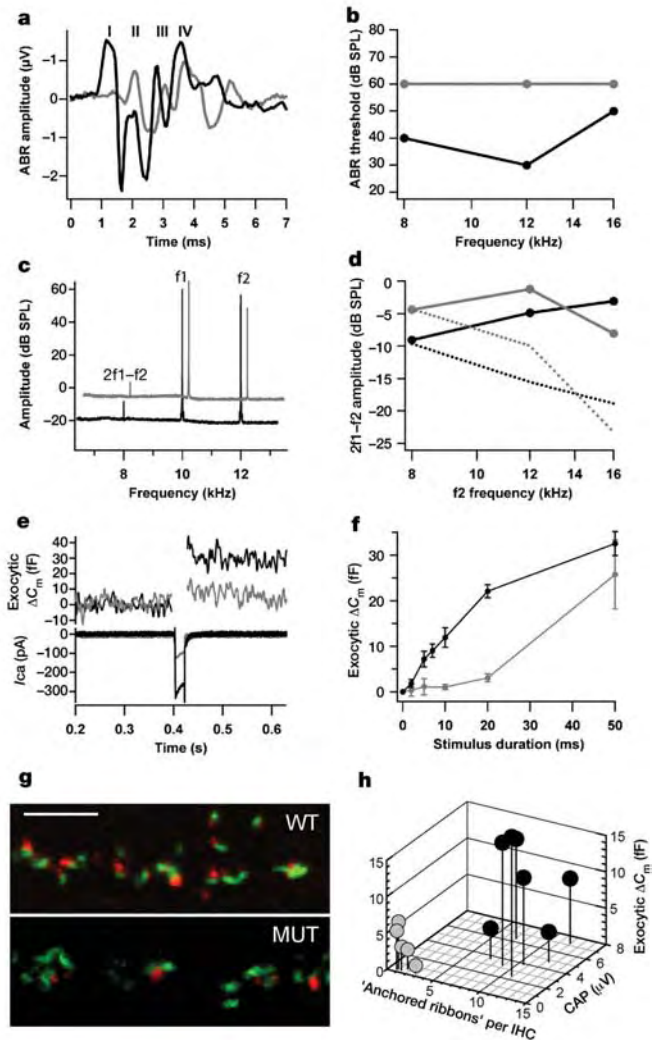


Figure 2 Synaptic ribbons are essential for hearing and fast exocytosis from hair cells. Analysis of a pair of mutant (grey) and wild-type (black) mice. **a**, Auditory brainstem response evoked by suprathreshold clicks (80 dB), peaks indicated with roman numbers. **b**, Auditory brainstem audiograms obtained by tone burst stimulation. **c**, Power spectrum of the microphone signal showing the 2f1–f2 distortion product otoacoustic emission (DPOAE), as well as the primary tones (f1 and f2), mutant data slightly shifted for better visibility. **d**, Distortion product audiogram. Markers show 2f1–f2 DPOAE levels, dotted lines show noise floor around 2f1–f2. **e**, Representative Ca^{2+} currents (I_{Ca}) and C_m changes recorded from wild-type and mutant IHCs (20-ms step depolarization). **f**, Kinetics of exocytosis constructed from ΔC_m of the same IHCs in response to depolarization of varying durations. Wild-type and mutant ΔC_m differed significantly for 5, 10 and 20 ms depolarizations. Errors are s.e.m. **g**, Immunostaining for RIBEYE (red) and GluR (green) of IHC afferent synapses. Single confocal sections through 4 neighbouring IHCs. Scale bar, 5 μm . **h**, The diagram relates the mean number of synapse-anchored ribbons per IHC, fast exocytosis (ΔC_m in response to 10-ms depolarization) and compound action potential (CAP) amplitude (approximated by auditory brainstem response peak I amplitude) for individual mutant (grey) and wild-type (black) mice (analysed as in **a–g**).

decreasing the extracellular Ca^{2+} concentration from 10 to 2 mM (asterisks in Fig. 3). Finally, we observed a reduction in fast exocytosis per unit of Ca^{2+} influx in mutant IHCs, although the Ca^{2+} sensitivity of release appeared to be unchanged when considering sustained exocytosis (Supplementary Fig. 3b). Mutant IHCs showed robust sustained exocytosis (Fig. 4) and endocytic membrane retrieval (Fig. 3b). Together, these observations indicate that the synaptic dysfunction of Bassoon mutant mice was primarily due to the reduction of synapse-anchored IHC ribbons, rather than to reduced Ca^{2+} influx or a general exocytic defect in the absence of functional Bassoon. This is consistent with a primary role for Bassoon in synaptic anchoring of ribbons¹⁰.

We then related membrane capacitance increments in large samples of ribbon-deficient and normal IHCs to their synapse morphology to dissect ribbon-dependent exocytosis and the mechanism for synaptic dysfunction. Exocytosis of 8-week-old wild-type IHCs was approximated by the sum of a small, fast secretory component and a large, slow component (Fig. 4). We interpret the fastest discernible component (time constant, $\tau \approx 9$ ms) as exocytosis of the readily releasable vesicle pool (RRP)²⁰. The total IHC RRP was estimated to ~ 18 fF or 640 synaptic vesicles by the model fit (conversion factor of 28 aF per synaptic vesicle, Supplementary Table 1). Mutant exocytosis was

subtracted to estimate the 'ribbon-dependent' RRP of ~ 15 fF or 530 synaptic vesicles (asterisks in Fig. 4a). Exocytosis of mutant IHCs could be better approximated by using a single component model. However, some residual fast exocytosis (~ 5 fF or 210 synaptic vesicles, conversion factor of 24 aF per synaptic vesicle, $\tau \approx 10$ ms) was evident. This is consistent with the morphological finding of docked vesicles at the ~ 11 ribbon-deficient synapses and ~ 1 ribbon-containing synapse per mutant IHC (Fig. 1f, Supplementary Table 1). Assuming that fast exocytosis exclusively represents synaptic transmitter release, each of the 10 ribbon-containing synapses of wild-type IHCs (9.8 ± 0.9 , $n = 7$ mice) on average holds a large RRP of 53–64 synaptic vesicles. This contrasts with a pool of ~ 14 synaptic vesicles (Supplementary Table 1) that can be released at comparable speed by the average ribbon-deficient synapse of Bassoon mutants, but does not support synchronous activation of SGNs.

It is not clear how many synaptic vesicles undergo exocytosis at a single afferent synapse after a brief acoustic stimulus *in vivo*. Postsynaptic recordings from immature IHC afferent synapses showed that transmitter release from a single vesicle is sufficient to depolarize the SGN to threshold. Nonetheless, synchronized release of several vesicles occurs frequently²¹. Our data suggest that synchronous auditory signalling indeed relies on precisely timed release of several synaptic vesicles at the mature afferent IHC synapse, which in turn requires the presence of the ribbon. Parallel and synchronized release of multiple synaptic vesicles at the IHC ribbon synapse is probably required to reduce jitter of postsynaptic action potentials. The achieved precision of stimulus coding enables auditory brainstem neurons to detect sub-millisecond interaural time differences²². In contrast to large auditory brainstem synapses, such as the calyx of Held, where parallel transmitter release at multiple active zones ensures temporally precise excitation of a large neuronal soma, parallel and precisely timed release of multiple vesicles at one ribbon-containing hair cell active zone excites a very small postsynaptic terminal³. Possible mechanisms include parallel fusion of docked synaptic vesicles and sequential fusion of more distal, ribbon-associated vesicles to the membrane of previously fused proximal vesicles²³. We favour the hypothesis that RRP exocytosis is dominated by fusion of docked vesicles^{1,4,5,24}, and that the ribbon positions synaptic vesicles close to

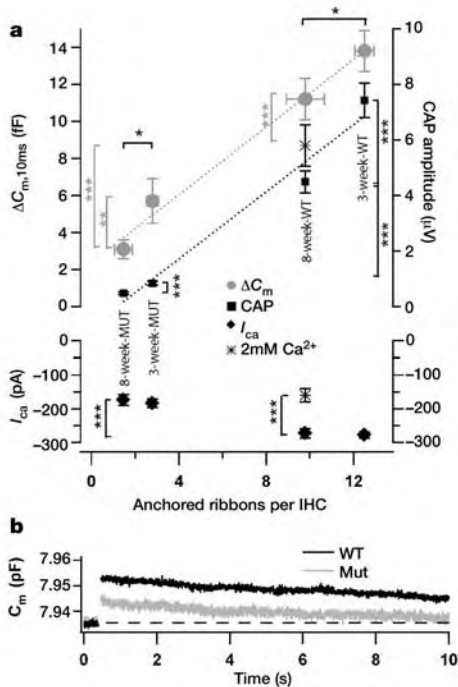


Figure 3 Fast exocytosis and compound action potential amplitude correlate with the number of anchored ribbons per IHCs during development. **a**, Mean ΔC_m (circles, upper panel) in response to 10-ms depolarization. Corresponding peak Ca^{2+} currents (diamonds, lower panel) and amplitude of auditory brainstem response peak I (squares, upper panel; responses to 80 dB clicks) of 3- and 8-week-old wild-type (WT) and mutant (Mut) mice were related to the numbers of synapse-anchored ribbons per IHC. ΔC_m and Ca^{2+} currents of 8-week-old wild-type IHCs were recorded at both 10 mM ($n = 25$ IHCs) and 2 mM extracellular Ca^{2+} (big asterisks, $n = 7$ IHCs). All other IHCs were recorded at 10 mM extracellular Ca^{2+} ($n = 24$ IHCs for 8-week-old mutant and $n = 12$ IHCs for both 3-week-old wild-type and mutant IHCs). Auditory brainstem results were obtained from 28 (wild-type) and 21 (mutant) ears of 8-week-old mice as well as from 14 (wild-type) and 6 (mutant) ears of 3-week-old mice. Counting of synapse-anchored ribbons was based on 80 wild-type and 60 mutant IHCs of 8-week-old mice, and on 78 wild-type and 94 mutant IHCs of 3-week-old mice. Dotted lines represent linear regressions (correlation coefficients > 0.95). All errors are s.e.m. **b**, Average exocytic and endocytic capacitance responses of 3-week-old wild-type and mutant IHCs to 20-ms depolarization.

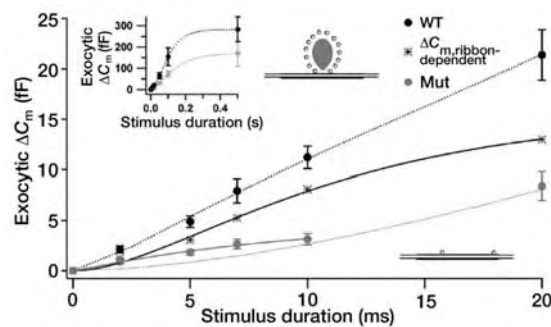


Figure 4 Dissection of ribbon-dependent hair cell exocytosis. Kinetics of exocytosis of wild-type (black circles, $n = 26$) and mutant (grey circles, $n = 24$) IHCs analysed as in Fig. 2f. Dotted lines are model fits to wild-type (black): $\Delta C_m = 18 \text{ fF} \times (1 - e^{-(t/9 \text{ ms})^{1.4}}) + 265 \text{ fF} \times (1 - e^{-(t/64 \text{ ms})^{2.8}}$ and mutant (grey): $\Delta C_m = 172 \text{ fF} \times (1 - e^{-(t/107 \text{ ms})^{1.7}}$. The inset shows the full range of data. The first 10 ms of mutant exocytosis were better approximated by $\Delta C_m = 5 \text{ fF} \times (1 - e^{-(t/11 \text{ ms})})$, grey solid line. The difference between wild-type and mutant ΔC_m is represented by asterisks, approximated by $\Delta C_m, \text{ribbon-dependent} = 15 \text{ fF} \times (1 - e^{-(t/8 \text{ ms})^2}$ (black solid line). Wild-type and mutant ΔC_m differed with $P < 0.01$ for stimuli below 50 ms, and with $P < 0.05$ for 50 ms, but were not significantly different for longer pulses. Cartoons illustrate typical two-dimensional sections of wild-type (top) and mutant (bottom) synapses. All errors are s.e.m.

Ca^{2+} channels to ensure efficient stimulus–secretion coupling. However, it was recently suggested that fast exocytosis in hair cells may be mediated by vesicles that are not docked at active zones^{25,26}.

To test the idea that synaptic vesicles tethered to the ribbon far from the presynaptic membrane might contribute to fast IHC exocytosis, we compared the number of readily releasable synaptic vesicles to the numbers of docked and total ribbon-associated synaptic vesicles. We analysed 42 ribbon-containing synapses of 8-week-old wild-type IHCs in single, ultrathin sections, and obtained upper estimates of docked and total ribbon-associated synaptic vesicles from ribbon surface calculations on the basis of 4Pi microscopy (Fig. 1k, Supplementary Table 1 and Supplementary Methods). We estimated between 125 (by electron microscopy) and 203 (by 4Pi microscopy) ribbon-associated synaptic vesicles and 16–30 docked synaptic vesicles; these values are comparable to estimates obtained from ribbons of goldfish bipolar neurons²⁷, but smaller than the vesicle load of spheroid bodies in frog saccular hair cells¹². Hence, the number of readily releasable synaptic vesicles (53–64) exceeded our estimate for the number of docked synaptic vesicles (16–30), suggesting that remote, tethered synaptic vesicles contribute to fast IHC exocytosis. However, it remains possible that we underestimated the number of docked synaptic vesicles owing to chemical fixation²⁸, and that the RRP entirely consists of docked synaptic vesicles in mouse IHCs.

Whereas the above results highlight the ribbon's function in synchronous synaptic transmission, the role of the ribbon in sustained exocytosis is less clear. The finding of robust, sustained exocytosis in mutant IHCs (Fig. 4, inset) is surprising in the light of the previously suggested 'conveyor belt' model of the ribbon^{1,4,5}. Our finding suggests that slow exocytosis occurs at ribbon-deficient synapses and that additional fusion outside the active zone also contributes^{18,24,29}. Indeed, the Bassoon mutants were not completely deaf, indicating that the remaining fast exocytosis and the presence of slow exocytosis support some residual auditory signalling. However, temporally precise sound coding is impaired in this mouse model of synaptic audiopathy. A defect of synchronous IHC synaptic transmission could present a pathomechanism of human auditory neuropathy, and explain the poor speech discrimination that is observed. □

Methods

Animals

Mice (3- and 8-week-old) with a targeted deletion for exons 4 and 5 of the *bassoon* gene³⁰ and their wild-type littermates were used for experiments. Animal handling followed national ethical guidelines.

Auditory brainstem responses and otoacoustic emissions

Mice were anaesthetized intraperitoneally with 100 mg kg⁻¹ ketamine and 2 mg kg⁻¹ xylazine. For auditory brainstem responses, tone bursts (8/12/16 kHz, 10 ms plateau, 1 ms rise/fall) or clicks of 0.03 ms were applied in the free field at 20 Hz. Intensities are presented as sound pressure level (dB root mean square for tone bursts, dB peak equivalent for clicks). The difference potential between vertex and mastoid intradermal needles was amplified (5×10^4 times), filtered (low-pass, 4 kHz; high-pass, 100 Hz) and sampled at a rate of 50 kHz for 20 ms, $2 \times 2,000$ times to obtain two mean auditory brainstem response traces for each sound intensity. Thresholds were estimated with 10 dB precision by means of visual inspection. Peak I amplitude was estimated as the difference between peak I and the subsequent negativity. For distortion product otoacoustic emissions (DPOAE), a 24-bit sound card together with the ED1/EC1 speaker system (Tucker-Davis) were used to generate two primary tones (f1 and f2) with a ratio of f2/f1 = 1.2 and a range for f2 of 8 to 16 kHz. Primary tones were coupled into the ear canal by a custom-made probe containing an MKE-2 microphone (Sennheiser) and adjusted to an intensity of 60 dB sound pressure level (SPL) at the ear drum. The microphone signal was amplified, sampled by the sound card and analysed by fast Fourier transformation. The peak amplitude of the 2f1–f2 DPOAE and the noise (average power around 2f1–f2) were measured using f1 = 60 dB SPL as reference.

Patch-clamp recording

IHCs from the apical coils of freshly dissected organs of Corti were patch-clamped in the perforated-patch configuration as described¹⁸. The standard pipette solution contained 150 mM caesium gluconate, 13 mM TEA-Cl, 10 mM CsOH-HEPES, 1 mM MgCl₂ and 250 $\mu\text{g ml}^{-1}$ amphotericin B. The standard bath solution contained 105 mM NaCl, 35 mM

TEA-Cl, 2.8 mM KCl, 10 mM CaCl₂, 1 mM MgCl₂, 10 mM NaOH-HEPES, 10 mM D-glucose. EPC-9 amplifiers (HEKA-electronics), controlled by Pulse software, were used to low-pass filter and sample currents at 20–40 kHz and at 2–5 kHz, respectively. We measured membrane capacitance increments (C_m) as previously described¹⁸. Cells were stimulated by depolarizations of different durations to peak Ca^{2+} current potential at intervals of 30 to 70 s. All currents were leak-corrected using a P/6-protocol. Passive electrical IHC properties are provided in Supplementary Table 2.

Immunostaining, confocal and 4Pi microscopy

The organs of Corti were fixed with 4% paraformaldehyde in 120 mM sodium phosphate buffer (1 h). After incubation of whole-mount preparations for 1 h in goat serum dilution buffer (16% normal goat serum, 450 mM NaCl, 0.3% Triton X-100, 20 mM phosphate buffer, pH 7.4) primary antibodies were applied overnight at 4 °C. The following antibodies were used: rabbit anti-Bassoon (Sap7P, diluted 1:1,000), guinea pig anti-Piccolo⁶ (1:1,000), RIBEYE rabbit antiserum (a gift from H. Brandstätter, 1:1,000–4,000), mouse IgG1 anti-CtBP2 (BD Biosciences, 1:100–200), rabbit anti-GluR2/3 (Chemicom, 1:1,000), rabbit anti-calbindin (Swants, 1:100). Secondary AlexaFluor-labelled antibodies (Molecular Probes, 1:200) were applied for 1 h. Confocal images were collected using a LSM 510 microscope (Carl Zeiss Jena) and analysed in LSM 5 Image Browser and Adobe Photoshop. Multifocal 4Pi-microscopy with water immersion lenses (NA 1.2; semiaperture angle 64°) at a two-photon excitation wavelength of 870 nm (average power ~1.5 mW for each of the 4Pi-foci) was done as described before^{14,15} (for details see Supplementary Methods). For 4Pi image restoration, nonlinear deconvolution with a measured point spread function was performed as described previously¹⁵.

Transmission electron microscopy

Cochleas were fixed by perfusing 2% glutaraldehyde in Millonig buffer (pH 7.4) through the oval and round windows, after gently opening the apex, and were then immersed in the same fixative overnight, rinsed in Millonig, and postfixed in 2% osmium tetroxide for 1 h. They were then rinsed in Millonig twice, dehydrated in a graded ethanol series (30–100%) and in propylene oxide, and embedded in Epon. The cochlear coils were separated, and the upper second turn was remounted for semi-thin and thin longitudinal sectioning, allowing access to several IHC synaptic poles on the same section. Ultrathin sections (70–100 nm; Leica-Reichert) were counter-stained with uranyl acetate and lead citrate, and observed with a Hitachi 7100 electron microscope.

Data analysis

Changes in C_m were estimated as the difference between the mean C_m after the end of the depolarization and the mean pre-pulse C_m (the initial 40 ms after the depolarization was omitted). Mean ΔC_m and Ca^{2+} current estimates present grand averages calculated from the mean estimates of individual IHCs. This avoided dominance of IHCs contributing more sweeps. All voltages were corrected for liquid junction potentials. Values were expressed as means $\pm$ s.e.m. and compared using Student's unpaired *t*-tests, with single, double and triple asterisks indicating $P < 0.05$, 0.02 and 0.01, respectively.

Received 22 November 2004; accepted 28 January 2005; doi:10.1038/nature03418.

- Fuchs, P. A., Glowatzki, E. & Moser, T. The afferent synapse of cochlear hair cells. *Curr. Opin. Neurobiol.* **13**, 452–458 (2003).
- Smith, C. A. & Sjostrand, F. S. A synaptic structure in the hair cells of the guinea pig cochlea. *J. Ultrastruct. Res.* **5**, 184–192 (1961).
- Lieberman, M. C. Single-neuron labeling in the cat auditory nerve. *Science* **216**, 1239–1241 (1982).
- Sterling, P. & Matthews, G. Structure and function of ribbon synapses. *Trends Neurosci.* **28**, 20–29 (2005).
- Lagnado, L. Ribbon synapses. *Curr. Biol.* **13**, R631 (2003).
- Schmitz, F., Konigstorfer, A. & Sudhof, T. C. RIBEYE, a component of synaptic ribbons: a protein's journey through evolution provides insight into synaptic ribbon function. *Neuron* **28**, 857–872 (2000).
- Dick, O. *et al.* Localization of the presynaptic cytomatrix protein Piccolo at ribbon and conventional synapses in the rat retina: comparison with Bassoon. *J. Comp. Neurol.* **439**, 224–234 (2001).
- tom Dieck, S. *et al.* Molecular dissection of the photoreceptor ribbon synapse: physical interaction of Bassoon and RIBEYE is essential for the assembly of the ribbon complex. *J. Cell Biol.* (in the press); published online 22 February 2005 (doi:10.1083/jcb.200408157).
- tom Dieck, S. *et al.* Bassoon, a novel zinc-finger CAG/glutamine-repeat protein selectively localized at the active zone of presynaptic nerve terminals. *J. Cell Biol.* **142**, 499–509 (1998).
- Dick, O. *et al.* The presynaptic active zone protein bassoon is essential for photoreceptor ribbon synapse formation in the retina. *Neuron* **37**, 775–786 (2003).
- Pailart, C., Li, J., Matthews, G. & Sterling, P. Endocytosis and vesicle recycling at a ribbon synapse. *J. Neurosci.* **23**, 4092–4099 (2003).
- Lenzi, D., Crum, J., Ellisman, M. H. & Roberts, W. M. Depolarization redistributes synaptic membrane and creates a gradient of vesicles on the synaptic body at a ribbon synapse. *Neuron* **36**, 649–659 (2002).
- Holt, M., Cooke, A., Wu, M. M. & Lagnado, L. Bulk membrane retrieval in the synaptic terminal of retinal bipolar cells. *J. Neurosci.* **23**, 1329–1339 (2003).
- Hell, S. & Stelzer, E. H. K. Properties of a 4Pi-confocal fluorescence microscope. *J. Opt. Soc. Am. A* **18**, 2159–2166 (1992).
- Egner, A., Jakobs, S. & Hell, S. W. Fast 100-nm resolution three-dimensional microscope reveals structural plasticity of mitochondria in live yeast. *Proc. Natl Acad. Sci. USA* **99**, 3370–3375 (2002).
- Kemp, D. T. Stimulated acoustic emissions from within the human auditory system. *J. Acoust. Soc. Am.* **64**, 1386–1391 (1978).
- Starr, A., Picton, T. W., Sininger, Y., Hood, L. J. & Berlin, C. I. Auditory neuropathy. *Brain* **119**, 741–753 (1996).
- Moser, T. & Beutner, D. Kinetics of exocytosis and endocytosis at the cochlear inner hair cell afferent synapse of the mouse. *Proc. Natl Acad. Sci. USA* **97**, 883–888 (2000).

19. Shneron, A., Devigne, C. & Pujol, R. Age-related changes in the C57BL/6J mouse cochlea. II. Ultrastructural findings. *Brain Res.* **254**, 77–88 (1981).
20. Neher, E. Vesicle pools and Ca^{2+} microdomains: new tools for understanding their roles in neurotransmitter release. *Neuron* **20**, 389–399 (1998).
21. Glowatzki, E. & Fuchs, P. A. Transmitter release at the hair cell ribbon synapse. *Nature Neurosci.* **5**, 147–154 (2002).
22. Trussell, L. O. Synaptic mechanisms for coding timing in auditory neurons. *Annu. Rev. Physiol.* **61**, 477–496 (1999).
23. Heidelberger, R., Heinemann, C., Neher, E. & Matthews, G. Calcium dependence of the rate of exocytosis in a synaptic terminal. *Nature* **371**, 513–515 (1994).
24. Zenisek, D., Steyer, J. A. & Almers, W. Transport, capture and exocytosis of single synaptic vesicles at active zones. *Nature* **406**, 849–854 (2000).
25. Edmonds, B. W., Gregory, F. D. & Schweizer, F. E. Evidence that fast exocytosis can be predominantly mediated by vesicles not docked at active zones in frog saccular hair cells. *J. Physiol. (Lond.)* **560**, 439–450 (2004).
26. Spassova, M. A. *et al.* Evidence that rapid vesicle replenishment of the synaptic ribbon mediates recovery from short-term adaptation at the hair cell afferent synapse. *J. Assoc. Res. Otolaryngol.* **5**, 376–390 (2004).
27. von Gersdorff, H., Vardi, E., Matthews, G. & Sterling, P. Evidence that vesicles on the synaptic ribbon of retinal bipolar neurons can be rapidly released. *Neuron* **16**, 1221–1227 (1996).
28. Smith, J. E. & Reese, T. S. Use of aldehyde fixatives to determine the rate of synaptic transmitter release. *J. Exp. Biol.* **89**, 19–29 (1980).
29. Beutner, D., Voets, T., Neher, E. & Moser, T. Calcium dependence of exocytosis and endocytosis at the cochlear inner hair cell afferent synapse. *Neuron* **29**, 681–690 (2001).
30. Altrock, W. D. *et al.* Functional inactivation of a fraction of excitatory synapses in mice deficient for the active zone protein bassoon. *Neuron* **37**, 787–800 (2003).

Supplementary Information accompanies the paper on www.nature.com/nature.

Acknowledgements We would like to thank A. Brandt and A. Schoenle for providing custom analysis software; S. Anderson and S. Lacas-Gervais for help in setting up auditory physiology and immunohistochemistry in the InnerEarLab, respectively; and J. H. Brandstaetter, F. Wolf and S. W. Hell for discussions and suggestions; members of the InnerEarLab for counting of spots and discussion; E. Neher, T. Sakaba, L. Lagnado, C. Kubisch, M. C. Liberman and E. Livesey for comments on the manuscript; F. Kirchhoff and M. Lenoir for help with confocal and electron microscopy, respectively; and M. Köppler, F. Tribillac and C. Cazeville for technical assistance. We would like to thank M. Eybalin for initial collaboration. This work was supported by grants from the Deutsche Forschungsgemeinschaft to T.M. and to E.D.G., by a Tandem-Project of the Max Planck Society (to E. Neher and T.M.), a Human Frontiers in Science Program grant to T.M., a grant from the Fonds der Chemischen Industrie to E.D.G., and a grant from Acouphènes-Languedoc-Roussillon to J.-L. Puel.

Competing interests statement The authors declare that they have no competing financial interests.

Correspondence and requests for materials should be addressed to T.M. (tmoser@gwdg.de).

Essential role of TRPC channels in the guidance of nerve growth cones by brain-derived neurotrophic factor

Yan Li^{1,2*}, Yi-Chang Jia^{1,2*}, Kai Cui^{1,2*}, Ning Li^{1,2}, Zai-Yu Zheng^{1,2}, Yi-zheng Wang¹ & Xiao-bing Yuan¹

¹Institute of Neuroscience, Shanghai Institutes for Biological Sciences, Chinese Academy of Sciences, Shanghai 200031, China

²Graduate School of the Chinese Academy of Sciences, Shanghai 200031, China

* These authors contributed equally to this work

Brain-derived neurotrophic factor (BDNF) is known to promote neuronal survival and differentiation¹ and to guide axon extension both *in vitro*^{2,3} and *in vivo*⁴. The BDNF-induced chemo-attraction of axonal growth cones requires Ca^{2+} signalling³, but how Ca^{2+} is regulated by BDNF at the growth cone remains largely unclear. Extracellular application of BDNF triggers membrane currents resembling those through TRPC (transient receptor potential canonical) channels in rat pontine neurons⁵ and in *Xenopus* spinal neurons⁶. Here, we report that in cultured cerebellar granule cells, TRPC channels contribute to the BDNF-induced elevation of Ca^{2+} at the growth cone and are

required for BDNF-induced chemo-attractive turning. Several members of the TRPC family are highly expressed in these neurons, and both Ca^{2+} elevation and growth-cone turning induced by BDNF are abolished by pharmacological inhibition of TRPC channels, overexpression of a dominant-negative form of TRPC3 or TRPC6, or downregulation of TRPC3 expression via short interfering RNA. Thus, TRPC channel activity is essential for nerve-growth-cone guidance by BDNF.

Growing axons are guided to their target cells by diffusible and substrate-bound guidance cues⁷, among which BDNF exerts both attractive and repulsive actions on growth cones^{2–4}. In the present study, we first examined the guidance action of BDNF on the growth cones of cultured rat cerebellar granule cells, the migration of which is stimulated by BDNF⁸. A microscopic gradient of BDNF was produced by pulsatile application of BDNF-containing solution with a micropipette near growth cones (see Methods). This BDNF gradient caused marked growth-cone turning towards the pipette within 30 min (Fig. 1a). Control experiments showed that application of vehicle solution (PBS) had no effect on growth-cone extension (Fig. 1b). The chemo-attractive effect of BDNF required activation of the high-affinity BDNF receptor TrkB, as it was abolished by bath application of K252a (Fig. 1c), an inhibitor of TrkB⁹. Furthermore, BDNF-induced growth-cone attraction was abolished in a Ca^{2+} -free medium (Fig. 1c), consistent with results in cultured *Xenopus* spinal neurons³. The attraction was also blocked by pre-incubation with BAPTA-AM (2 μM), a membrane-permeable Ca^{2+} buffer, or with thapsigargin (1 μM), which depletes intracellular Ca^{2+} stores (Fig. 1c). Therefore, Ca^{2+} influx through plasma membrane Ca^{2+} channels and Ca^{2+} release from internal stores are both required for growth-cone turning induced by BDNF.

It is known that netrin-1-induced growth-cone attraction requires Ca^{2+} influx through L-type voltage-dependent Ca^{2+} channels (VDCCs) and Ca^{2+} release from internal stores¹⁰. However, we found that BDNF-induced growth-cone attraction in rat cerebellar granule cells was unaffected by bath application of nifedipine (5 μM) and ω -conotoxin MVIIC (5 μM)—specific blockers of L- and P/Q-type VDCCs, respectively—or by applying ryanodine at a high concentration (100 μM , ref. 10) that blocks ryanodine-sensitive channels of internal stores (Fig. 1c). Thus, Ca^{2+} influx through VDCCs and internal Ca^{2+} release through ryanodine channels are unlikely to contribute to BDNF-induced growth-cone turning. Rapid neuronal depolarization through the activation of a saxitoxin-sensitive Na^{+} channel ($\text{Na}_v1.9$) may be evoked by BDNF¹¹, but neither saxitoxin (10 nM) nor tetrodotoxin (50 nM, data not shown) had an effect on BDNF-induced growth-cone turning (Fig. 1c), suggesting that Na^{+} -channel-dependent depolarization is not involved.

Because Ca^{2+} -permeant channels other than VDCCs appear to mediate the Ca^{2+} influx required for BDNF chemo-attraction, we examined the involvement of Ca^{2+} -permeant TRPC channels. When granule cells were incubated in SKF-96365 (3 μM), a relatively nonspecific inhibitor of store-operated Ca^{2+} entry and of TRP channels^{5,12}, BDNF-induced growth-cone turning was completely abolished (Fig. 1a, b), suggesting that Ca^{2+} entry through TRP channels may be involved. The blocking effect of SKF-96365 is specific to BDNF, because this drug had no effect on growth-cone turning induced by a glutamate gradient (Fig. 1b, 0.5 mM). The latter turning was abolished instead by antagonists of NMDA (*N*-methyl-D-aspartate) receptors, AMPA (α -amino-3-hydroxy-5-methyl-4-isoxazole propionic acid) receptors and L-type VDCCs, but not by inhibitors of metabotropic glutamate receptors (Supplementary Fig. 1). Similarly, the repulsive turning triggered by a gradient of stromal cell-derived factor 1 (SDF-1, 20 $\mu\text{g ml}^{-1}$), a chemokine known to repel the growth cones of these granule cells¹³, was not affected by SKF-96365 (Fig. 1b).

Additional studies were carried out to examine signalling events triggered by BDNF, using pharmacological treatments (Fig. 2c).

First, BDNF-induced chemo-attraction required the activation of phospholipase C γ (PLC- γ), a downstream effector of TrkB¹. A specific PLC inhibitor, U73122 (20 nM), abolished the chemo-attractive effect of BDNF (Fig. 2a), whereas its inactive derivative U7343 (20 nM) had no effect. Moreover, bath application of phosphatidylinositol-3-OH kinase (PI(3)K) inhibitors wortmannin (50 nM) or LY294002 (10 μ M) abolished BDNF-induced turning, whereas inhibiting MAP kinase with U0126 (1 μ M) and PD98059 (10 μ M) had no effect (Fig. 2a). These results showed that both PLC- γ and PI(3)K are required for BDNF-induced turning, consistent with data for *Xenopus* spinal neurons¹⁴.

We investigated whether PLC activation by itself is sufficient to trigger growth-cone turning. We found that a gradient of the PLC activator *m*-3M3FBS (27 mM in the pipette) caused marked chemo-attractive growth-cone turning, and this turning was abolished by removing extracellular Ca²⁺ and by SKF-96365 treatment (Fig. 2b), suggesting a requirement of TRPC channels. Together with results described earlier, we conclude that PLC signalling is necessary for BDNF-induced turning and that it is sufficient by itself to trigger the turning response. Furthermore, TRPC channels appeared to be downstream of PLC in mediating the Ca²⁺ influx required for the turning response.

A direct downstream effector of PLC is inositol-1,4,5-trisphosphate (InsP₃), which binds to the InsP₃ receptor (InsP₃R) and triggers internal Ca²⁺ release^{15,16}. The TRPC3 channels expressed in HEK-293 cells interact with InsP₃Rs (ref. 17), and activation of InsP₃ triggers BDNF-induced TRPC3 currents in cultured pontine neurons⁵. Blocking InsP₃R with xestospongine C (2 μ M) or 2APB

(100 μ M) abolished BDNF-induced growth-cone attraction in these granule cells (Fig. 2a). We have also examined the effects of perturbing diacylglycerol (DAG)–protein kinase C (PKC) signalling, which is also triggered by PLC. In contrast to InsP₃R inhibition, two specific PKC inhibitors, GF109203X (50 nM) and Go6976 (20 nM), had no effect on growth-cone attraction induced by a gradient of either BDNF (Fig. 2a) or *m*-3M3FBS (Fig. 2b). A gradient of 1-oleoyl-2-acetyl-*sn*-glycerol (OAG, 50 mM in the pipette), a membrane-permeable DAG analogue¹⁸, also had no turning effects (Fig. 2b). Thus, BDNF-induced attraction depends on InsP₃R activation, but not DAG–PKC signalling.

The above results suggest that BDNF acts on cerebellar granule cells by activating PLC- γ , leading to subsequent InsP₃R activation and Ca²⁺ release from internal stores (Fig. 2c). Furthermore, PLC- γ may also activate TRPC channels⁵, providing the Ca²⁺ influx that further elevates intracellular Ca²⁺ concentration ([Ca²⁺]_i). Fluorescence ratio imaging (see Methods) of [Ca²⁺]_i at growth cones showed a significant elevation of [Ca²⁺]_i in growth cones after BDNF application (at 2 μ g ml⁻¹ in the pipette; Fig. 3a, b), and this elevation was suppressed by thapsigargin (1 μ M; Fig. 2a, b) or by removing extracellular Ca²⁺ (Fig. 3c). Thus, BDNF-induced elevation of [Ca²⁺]_i depends on both internal Ca²⁺ release and Ca²⁺ influx. Notably, treatment with SKF-96365 at the same concentration that abolished growth-cone turning (Fig. 1a, b) markedly reduced BDNF-induced elevation of [Ca²⁺]_i (Fig. 3b, c) but had no effect on that induced by glutamate. Moreover, bath application of nifedipine, ryanodine, or saxitoxin failed to affect BDNF-induced elevation of [Ca²⁺]_i, consistent with that found

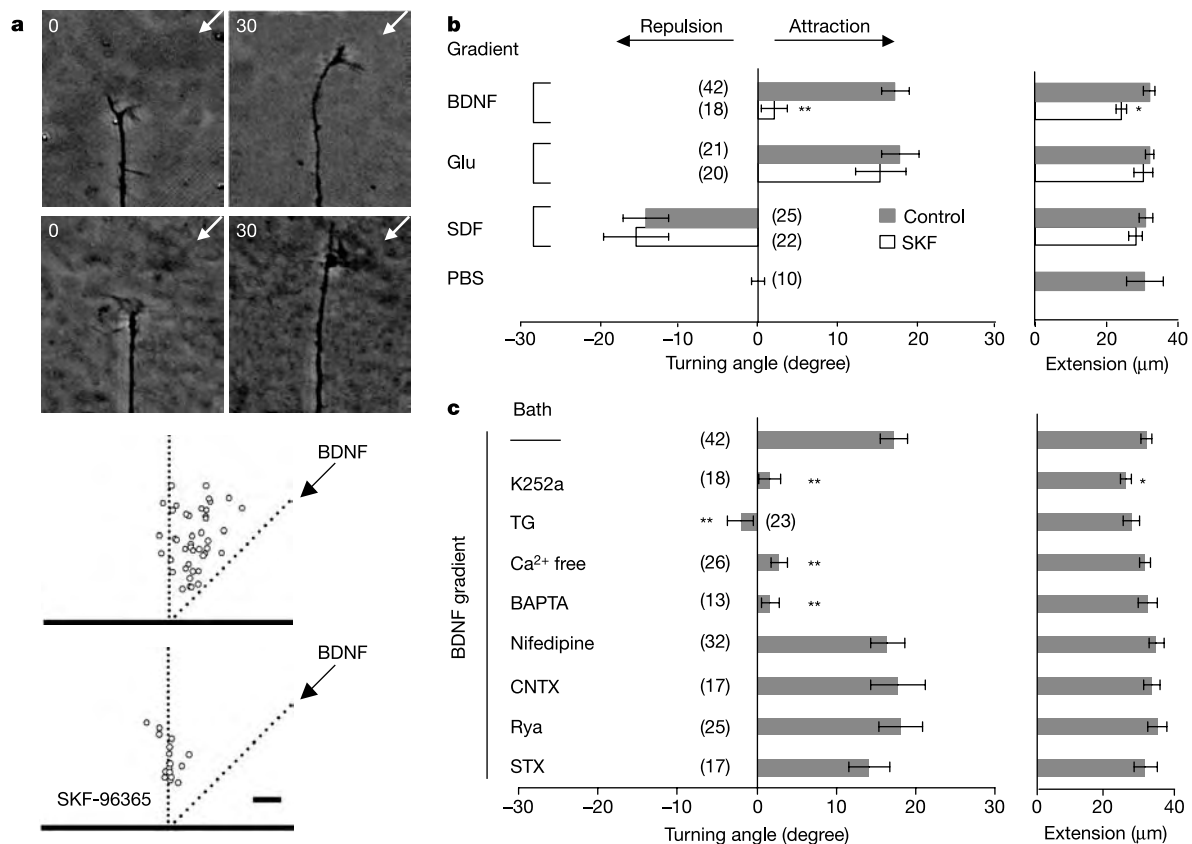


Figure 1 BDNF-induced growth-cone turning. **a**, Images of growth cones of cultured cerebellar granule cells before and after 30 min exposure to a BDNF gradient (arrow) in the absence or presence of SKF-96365. Plots depict the final positions of growth cones. The scale bar (10 μ m) applies to all panels in **a**. **b**, Average turning angles and average length of neurite extension after 30 min in a gradient of BDNF, glutamate, or SDF-1, in the absence or presence of SKF-96365, K252a (200 nM), thapsigargin (TG), BAPTA-AM,

nifedipine, ω -conotoxin MVIC (CNTX), ryanodine (Rya) or saxitoxin (STX). Numbers in parentheses refer to the total number of growth cones examined. For all figures, error bars indicate standard error of the mean (s.e.m.). For all growth-cone turning results, data marked with asterisks are significantly different from the control ($P < 0.01$, Kolmogorov–Smirnov test).

for growth-cone turning. That TRPC activation is downstream of PLC- γ and InsP₃R activation is further confirmed by the finding that either U73122 or xestospongin C significantly reduces BDNF-induced elevation of $[Ca^{2+}]_i$ (Fig. 3c). Taken together, these results support the model (Fig. 2c) in which BDNF triggers Ca^{2+} release from internal stores through activation of the PLC- γ -InsP₃ pathway, which in turn activates TRPC channels and allows the enhanced elevation of $[Ca^{2+}]_i$ that is required for triggering the attractive turning. Whether there is a direct interaction between InsP₃Rs and TRPC channels, as found in HEK-293 cells¹⁷, remains to be determined.

To confirm further that TRPC channels are involved in BDNF-induced chemo-attraction, we examined TRPC expression in cultured cerebellar granule cells. As shown in Fig. 4a, polymerase chain reaction with reverse transcription (RT-PCR) experiments showed that messenger RNAs of TRPC1, TRPC3 and TRPC6 are

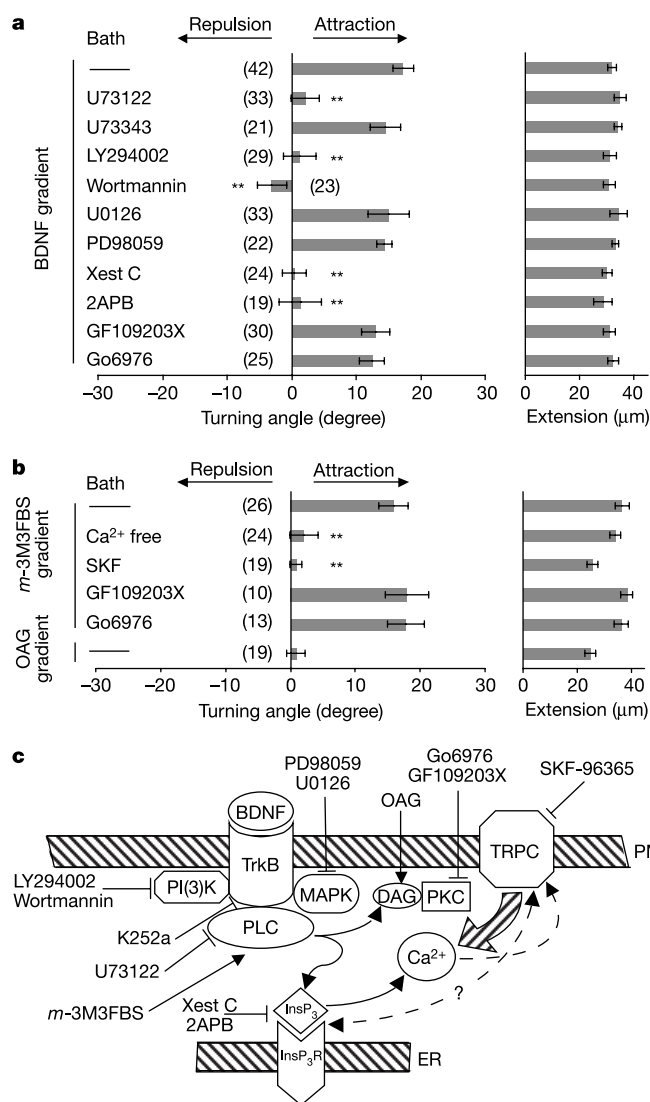


Figure 2 PLC mediates growth-cone attraction by BDNF. **a**, The average turning angles and neurite extension after 30 min exposure to a BDNF gradient in medium containing U73122, U73343, LY294002, wortmannin, U0126, PD98059, xestospongin C (Xest C), 2APB, GF109203X or Go6976. **b**, Growth-cone turning in a gradient of the PLC activator *m*-3M3FBS or a diacylglycerol analogue OAG, in the absence or presence of SKF-96365, GF109203X and Go6976. **c**, Schematic diagram depicting a working model for BDNF-triggered signalling that leads to elevation of $[Ca^{2+}]_i$, with target sites for drugs used. ER, endoplasmic reticulum; PM, plasma membrane.

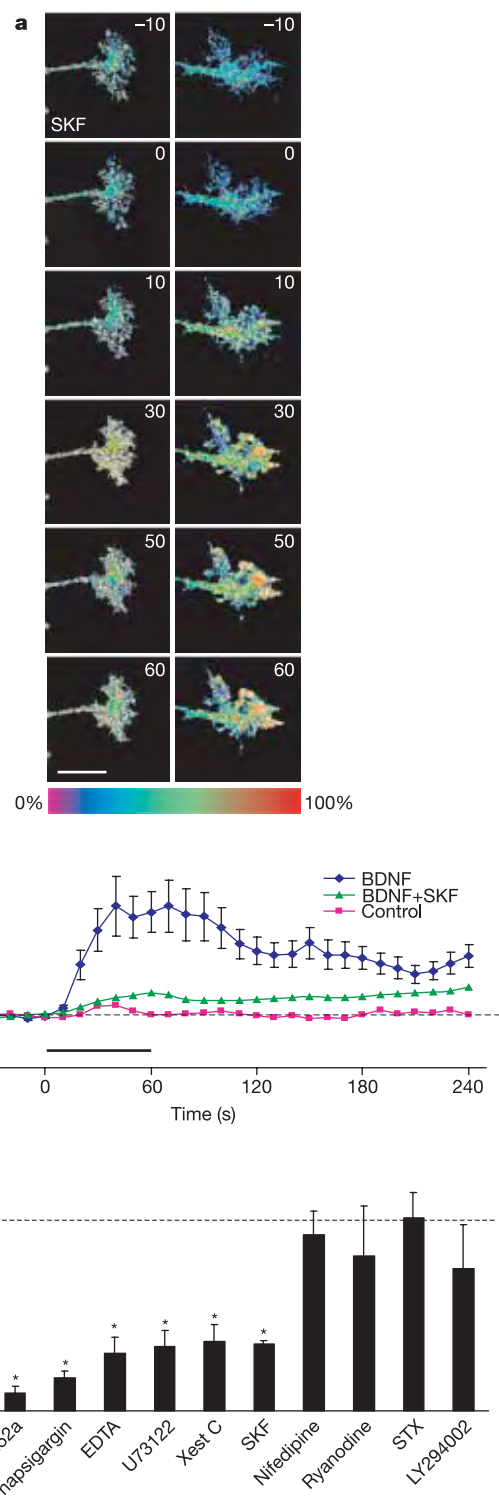


Figure 3 BDNF-induced Ca^{2+} elevation in the growth cone. **a**, Images of a growth cone loaded with Fluo-4 AM and Fura red AM before and after the onset of BDNF application (at time 0), with the ratio of Fluo-4 AM and Fura red AM fluorescence coded in pseudo-colour in a linear scale (0–100%). The panels depict growth-cone Ca^{2+} levels in the absence or presence of SKF-96365, respectively. Numbers indicate the time in seconds. Scale bar, 5 μ m. **b**, Changes in the growth-cone Ca^{2+} level after BDNF application (bar), depicted as the percentage change ($\Delta R/R$)100% in the ratio (*R*) of Fluo-4/Fura red AM, normalized to the baseline ratio for the same growth cone. **c**, Summary of changes in fluorescence ratio induced by BDNF in the absence or presence of various drugs.

highly expressed in both cerebellar tissue and granule cell cultures. The expression of TRPC1, TRPC3 and TRPC6 proteins was also confirmed by western blotting (Fig. 4b). To examine further which TRPC protein is required for BDNF-induced attraction, we down-regulated the expression of these proteins in cultured granule cells by specific short interfering RNAs (siRNAs) against TRPC1 and TRPC3. For each TRPC protein, at least four siRNA sequences were designed and the one with the highest efficiency and specificity was selected (Supplementary Table 1). A siRNA against TRPC3 caused marked reduction in the expression of TRPC3 protein, without influencing the expression of TRPC1 or TRPC6 (Fig. 4b). Similarly, a siRNA against TRPC1 caused reduction of TRPC1 expression without influencing TRPC3 or TRPC6 expression. Furthermore, we found that siRNA against TRPC3 abolished BDNF-induced growth-cone attraction, whereas that against TRPC1 or the nonsense control RNA had no effect (Fig. 4c). The effect of siRNA is specific to BDNF-induced turning, because the glutamate-induced growth-cone attraction was not affected (Supplementary Fig. 2a).

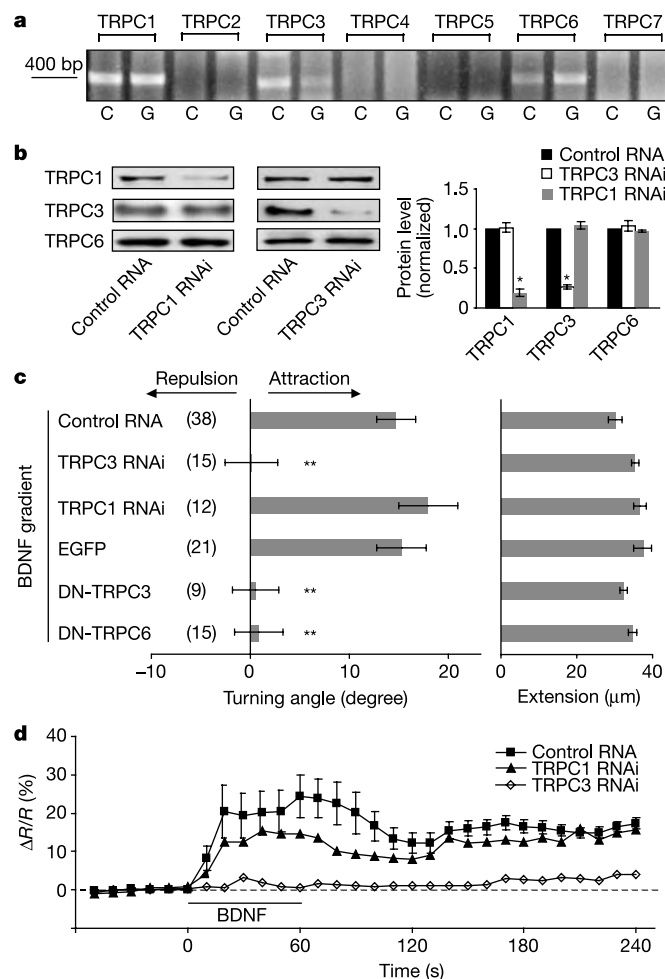


Figure 4 Downregulation of TRPC proteins affects BDNF-induced growth-cone attraction and Ca^{2+} elevation. **a**, Expression of TRPC mRNAs in cerebellar tissue (C) and cultures of granule cells (G), as shown by RT-PCR. **b**, Representative western blots and quantitative analysis showing specific downregulation of TRPC1 and TRPC3 by corresponding siRNA. The relative levels of TRPC expression were determined by densitometry measurements. Data (mean $\pm$ s.e.m., $n = 3$) are normalized by the nonsense control group. **c**, Effects of the BDNF gradient on growth-cone turning in neurons transfected with specific siRNAs and control RNA. DN, dominant negative. **d**, Changes in the growth cone Ca^{2+} level ($\Delta F/F$) over time for neurons transfected with siRNAs and control RNA.

The essential role of TRPC channels in BDNF-induced chemo-attraction was further confirmed by expressing constructs of green fluorescent protein (GFP)-tagged dominant-negative (DN)-TRPC3 or DN-TRPC6 (refs 19, 20). Cultured cerebellar granule cells were transfected with either one of these constructs or with a GFP control construct using electroporation. Cells with detectable GFP fluorescence were used for assaying growth-cone turning. Overexpression of either DN-TRPC3 or DN-TRPC6 in these granule cells abolished the BDNF-induced attraction (Fig. 4c), indicating the requirement for both TRPC3 and TRPC6. Furthermore, BDNF-induced elevation of $[Ca^{2+}]_i$ was also significantly reduced in neurons treated with the siRNA against TRPC3 but not with that against TRPC1 (Fig. 4d). In contrast, none of the siRNAs against TRPC1 or TRPC3 blocked glutamate-induced elevation of $[Ca^{2+}]_i$ (Supplementary Fig. 2b). Thus, both TRPC3 and TRPC6, but not TRPC1, are specifically required for BDNF-induced growth-cone chemo-attraction in these granule cells.

Elevation of $[Ca^{2+}]_i$ is essential for transducing the guidance signal of several putative guidance factors, but the cause of $[Ca^{2+}]_i$ elevation seems to differ among various factors. Elevation of $[Ca^{2+}]_i$ can result from Ca^{2+} influx through VDCCs in response to membrane depolarization or from opening of ligand-gated Ca^{2+} channels; for example, ionotropic glutamate receptors. Internal release of Ca^{2+} triggered by $InsP_3$ or by Ca^{2+} itself may also contribute to elevation of $[Ca^{2+}]_i$. Here, we showed that an $InsP_3$ -dependent Ca^{2+} influx via TRPC channels contributes to elevation of $[Ca^{2+}]_i$ in cultured cerebellar granule cells. Our results suggest that BDNF activates the PLC- γ - $InsP_3$ pathway, which further induces TRPC currents, resulting in Ca^{2+} influx and subsequent growth-cone attraction.

Among various TRPC proteins, TRPC3, TRPC6 and TRPC7 belong to the same subfamily^{22,23} and can form functional heteromeric or homomeric channels^{19,24}. In cultured hippocampal neurons, expression of DN-TRPC5 elevates neurite extension, whereas overexpression of wild-type TRPC5 inhibits neurite growth²⁵. However, we found that downregulation of TRPC3 by siRNA did not significantly affect neurite growth (data not shown). Owing to longer, mean channel open times²⁶⁻²⁸, TRPC5-containing channels may allow a larger Ca^{2+} influx than TRPC3/6 channels, leading to neurite growth inhibition, whereas a modest level of Ca^{2+} influx through TRPC3/6 channels is sufficient to trigger growth-cone turning. By allowing different patterns of Ca^{2+} influx, diverse TRPC channels may thus carry out distinct regulatory functions at the growth cone. Because many extracellular factors activate PLC²⁹, PLC-dependent TRP channel activation may also mediate Ca^{2+} signalling for other guidance factors. □

Methods

Cell culture

Cerebellum tissues of P0 Sprague-Dawley rats were incubated in CMF-PBS (calcium- and magnesium-free PBS) containing 0.1% trypsin for 10 min (37 °C), followed by trituration. Dissociated cells were collected by centrifugation, resuspended, and plated on coverslips coated with Matrigel (BD Biosciences). Cells were used for the turning assay or Ca^{2+} imaging 20–40 h after plating.

Plasmids and siRNA transfection

DN-TRPC3 was constructed by cloning one hTRPC3 fragment (1–302 amino acids of hTRPC3) into pEGFP-N1 vector (Clontech). This fragment was obtained by PCR amplification from the full-length hTRPC3 (a gift from J. W. Putney). DN-TRPC6 (with three highly conserved amino acids, L678 to W680, within the putative pore region of hTRPC6 replaced) was from T. Gudermann. For complementary DNA and siRNA transfection, we used the rat Neuron Nucleofector Kit (Amaxa) according to the manufacturer's instructions. Briefly, dissociated cells were resuspended in transfection medium, mixed with plasmids (6 μg) or siRNAs (200 nM), and electroporated using the fixed programme (O-03) for optimal neuronal transfection. Cells were then quickly centrifuged, resuspended and plated. The knockdown efficiency was determined by western blotting of extracts from transfected cultures.

Growth-cone turning assay

Microscopic gradients of chemicals were produced as described previously¹³. To assay

growth-cone turning, a micropipette was placed 75 μm away from the growth-cone centre and at an angle of 45° with respect to the initial direction of the neurite extension. Previous analysis has shown that, under standard assay conditions, the average extracellular concentration of BDNF at the growth cone is approximately 1/1,000 of that in the micropipette⁴⁰. Only growth cones with net extension >10 μm during the 30-min turning assay were analysed. L-15 medium (GIBCO) was used during the turning assay.

Ca²⁺ imaging

Neurons were rinsed before being loaded with Fluo-4 AM and Fura red AM (2 μM) with 0.1% dimethyl sulphoxide (DMSO) in NeuralBasal (GIBCO) medium for 30 min at room temperature, and then incubated for another 30 min after being rinsed in L-15 twice. Imaging was performed on a Zeiss LSM 510 confocal microscope with a $\times 60$ oil objective (Zeiss Plan-Neofluar, NA, 1.3). Cells were excited with a laser at 488 nm, and the intensity of the fluorescence between 505–550 nm was measured as the Fluo-4 signal, while fluorescence >635 nm was simultaneously detected as the Fura red signal. Images were acquired at 2-s intervals. The change in Ca²⁺ signal at the growth cone was determined by the percentage change in the ratio (R) of Fluo-4 to Fura red fluorescence (($\Delta R/R$)100%) in the entire growth cone, normalized by the baseline ratio measured at the same growth cone before BDNF or glutamate treatment. The BDNF and glutamate gradients were applied by pulsatile application of a high concentration solution (2 $\mu\text{g ml}^{-1}$ BDNF, 0.5 mM glutamate) from a micropipette at about 70 μm from the growth cone, using pulsing conditions similar to that used in the turning assay.

RT-PCR and immunoblotting

For RT-PCR, total RNA was isolated from cerebellum tissue and cultures of cerebellar granule neurons with Trizol reagent (Invitrogen). About 5 μg of total RNA was converted to cDNA (RevertAid First Strand cDNA Synthesis Kit, MBI Fermentas), and 1/20 of the reaction was used in 50- μl PCR reactions (Perkin Elmer). Conditions for the PCR reaction were: 94 °C, 30 s, 60 °C, 30 s, 72 °C, 45 s for 30 cycles, with a 10-s 72 °C final extension. Primers used are shown in Supplementary Table 1. End reaction products were visualized on ethidium-bromide-stained 1.0% agarose gels. To detect the expression level of TRPC proteins, cultured cerebellar granule cells were collected by lysis buffer (0.1% SDS, 1% nonidet P-40, 1% glycerol, 50 mM HEPES, pH 7.4, 2 mM EDTA and 100 mM NaCl). Proteins were separated by 10% SDS-polyacrylamide gel electrophoresis and transferred onto PVDF membrane. Membranes were blocked over night at 4 °C in 5% BSA and incubated with a polyclonal antibody (1:500, Alamone Labs) for 24 h at 4 °C, rinsed and incubated for 1 h with a horseradish-peroxidase-conjugated goat antibody against rabbit IgG (1:10,000; Biorad). Chemiluminescence detection was performed with the ECL kit (Pierce). The specificity of TRPC antibodies was tested using western blotting (Supplementary Fig. 3).

Sources and preparation of reagents

BDNF was purchased from R&D Systems; ryanodine and thapsigargin from Alomone Labs; saxitoxin from Sigma-Aldrich; and Fluo-4 AM and Fura red AM from Molecular Probes. The siRNAs were synthesized (Supplementary Table 1) by Shanghai GeneChem. All other chemicals were purchased from Calbiochem. BDNF was dissolved in water and prepared in 50 $\mu\text{g ml}^{-1}$ stock solution. Other chemicals were dissolved in DMSO in $\times 1,000$ stock solution, and diluted before use. For pharmacological treatments, chemicals were added to the culture medium at 30 min before the start and were present throughout the experiments.

Received 19 December 2004; accepted 21 February 2005; doi:10.1038/nature03477.
Published online 9 March 2005.

- Huang, E. J. & Reichardt, L. F. Trk receptors: roles in neuronal signal transduction. *Annu. Rev. Biochem.* **72**, 609–642 (2003).
- Yuan, X. B. *et al.* Signalling and crosstalk of Rho GTPases in mediating axon guidance. *Nature Cell Biol.* **5**, 38–45 (2003).
- Song, H. J., Ming, G. L. & Poo, M. cAMP-induced switching in turning direction of nerve growth cones. *Nature* **388**, 275–279 (1997).
- Ernfors, P., Lee, K. F. & Jaenisch, R. Mice lacking brain-derived neurotrophic factor develop with sensory deficits. *Nature* **368**, 147–150 (1994).
- Li, H. S., Xu, X. Z. & Montell, C. Activation of a TRPC3-dependent cation current through the neurotrophin BDNF. *Neuron* **24**, 261–273 (1999).
- Gordon, X. W. & Poo, M. Requirement of TRPC channels in netrin-1-induced chemotropic turning of nerve growth cones. *Nature* doi:10.1038/nature03478 (this issue).
- Tessier-Lavigne, M. & Goodman, C. S. The molecular biology of axon guidance. *Science* **274**, 1123–1133 (1996).
- Borghesani, P. R. *et al.* BDNF stimulates migration of cerebellar granule cells. *Development* **129**, 1435–1442 (2002).
- Tapley, P., Lamballe, E. & Barbacid, M. K252a is a selective inhibitor of the tyrosine protein kinase activity of the trk family of oncogenes and neurotrophin receptors. *Oncogene* **7**, 371–381 (1992).
- Hong, K., Nishiyama, M., Henley, J., Tessier-Lavigne, M. & Poo, M. Calcium signalling in the guidance of nerve growth by netrin-1. *Nature* **403**, 93–98 (2000).
- Blum, R., Kafitz, K. W. & Konnerth, A. Neurotrophin-evoked depolarization requires the sodium channel Na_v1.9. *Nature* **419**, 687–693 (2002).
- Merritt, J. E. *et al.* SKF 96365, a novel inhibitor of receptor-mediated calcium entry. *Biochem. J.* **271**, 515–522 (1990).
- Xiang, Y. *et al.* Nerve growth cone guidance mediated by G protein-coupled receptors. *Nature Neurosci.* **5**, 843–848 (2002).
- Ming, G. *et al.* Phospholipase C- γ and phosphoinositide 3-kinase mediate cytoplasmic signaling in nerve growth cone guidance. *Neuron* **23**, 139–148 (1999).
- Prentki, M. *et al.* Rapid mobilization of Ca²⁺ from rat insulinoma microsomes by inositol-1,4,5-trisphosphate. *Nature* **309**, 562–564 (1984).

- Burgess, G. M. *et al.* The second messenger linking receptor activation to internal Ca release in liver. *Nature* **309**, 63–66 (1984).
- Kiselyov, K. *et al.* Functional interaction between InsP₃ receptors and store-operated Htrp3 channels. *Nature* **396**, 478–482 (1998).
- Hofmann, T. *et al.* Direct activation of human TRPC6 and TRPC3 channels by diacylglycerol. *Nature* **397**, 259–263 (1999).
- Hofmann, T., Schaefer, M., Schultz, G. & Gudermann, T. Subunit composition of mammalian transient receptor potential channels in living cells. *Proc. Natl Acad. Sci. USA* **99**, 7461–7466 (2002).
- Groschner, K. *et al.* Trp proteins form store-operated cation channels in human vascular endothelial cells. *FEBS Lett.* **437**, 101–106 (1998).
- Henley, J. R., Huang, K. H., Wang, D. & Poo, M. M. Calcium mediates bidirectional growth cone turning induced by myelin-associated glycoprotein. *Neuron* **44**, 909–916 (2004).
- Trebak, M., Vazquez, G., Bird, G. S. & Putney, J. W. Jr. The TRPC3/6/7 subfamily of cation channels. *Cell Calcium* **33**, 451–461 (2003).
- Clapham, D. E., Runnels, L. W. & Strubing, C. The TRP ion channel family. *Nature Rev. Neurosci.* **2**, 387–396 (2001).
- Goel, M., Sinkins, W. G. & Schilling, W. P. Selective association of TRPC channel subunits in rat brain synaptosomes. *J. Biol. Chem.* **277**, 48303–48310 (2002).
- Greka, A., Navarro, B., Oancea, E., Duggan, A. & Clapham, D. E. TRPC5 is a regulator of hippocampal neurite length and growth cone morphology. *Nature Neurosci.* **6**, 837–845 (2003).
- Yamada, H. *et al.* Spontaneous single-channel activity of neuronal TRP5 channel recombinantly expressed in HEK293 cells. *Neurosci. Lett.* **285**, 111–114 (2000).
- Jung, S. *et al.* Lanthanides potentiate TRPC5 currents by an action at extracellular sites close to the pore mouth. *J. Biol. Chem.* **278**, 3562–3571 (2003).
- Zitt, C. *et al.* Expression of TRPC3 in Chinese hamster ovary cells results in calcium-activated cation currents not related to store depletion. *J. Cell Biol.* **138**, 1333–1341 (1997).
- Rebecchi, M. J. & Pentyala, S. N. Structure, function, and control of phosphoinositide-specific phospholipase C. *Physiol. Rev.* **80**, 1291–1335 (2000).
- Lohof, A. M., Quillan, M., Dan, Y. & Poo, M. M. Asymmetric modulation of cytosolic cAMP activity induces growth cone turning. *J. Neurosci.* **12**, 1253–1261 (1992).

Supplementary Information accompanies the paper on www.nature.com/nature.

Acknowledgements We thank T. Gudermann for providing the cDNA of DN-TRPC6; J. W. Putney for full-length hTRPC3; K. Groschner for DN-TRPC3; Q. Hu for technical support with confocal microscopy; J.-Y. Yu for cell culture; and Y.-J. Zhang for some biochemical assays. This work was supported by grants from Major State Basic Research Program and National Natural Science Foundation of China. X.-b.Y. was also supported by the Shanghai city government.

Competing interests statement The authors declare that they have no competing financial interests.

Correspondence and requests for materials should be addressed to X.-b.Y. (yuanxb@ion.ac.cn) or Y.-z.W. (yzwang@ion.ac.cn).

Requirement of TRPC channels in netrin-1-induced chemotropic turning of nerve growth cones

Gordon X. Wang & Mu-ming Poo

Division of Neurobiology, Department of Molecular and Cell Biology, Helen Wills Neuroscience Institute, University of California, Berkeley, California 94720, USA

Ion channels formed by the TRP (transient receptor potential) superfamily of proteins act as sensors for temperature, osmolarity, mechanical stress and taste^{1,2}. The growth cones of developing axons are responsible for sensing extracellular guidance factors, many of which trigger Ca²⁺ influx at the growth cone^{3,4}; however, the identity of the ion channels involved remains to be clarified. Here, we report that TRP-like channel activity exists in the growth cones of cultured *Xenopus* neurons and can be modulated by exposure to netrin-1 and brain-derived neurotrophic factor, two chemoattractants for axon guidance. Whole-cell recording from growth cones showed that netrin-1 induced a membrane depolarization, part of which remained after all major voltage-dependent channels were blocked. Furthermore, the membrane depolarization was sensitive to blockers of TRP channels. Pharmacological blockade of putative TRP currents or down-regulation of *Xenopus* TRP-1 (xTRPC1) expression with a specific

morpholino oligonucleotide abolished the growth-cone turning and Ca^{2+} elevation induced by a netrin-1 gradient. Thus, TRPC currents reflect early events in the growth cone's detection of some extracellular guidance signals, resulting in membrane depolarization and cytoplasmic Ca^{2+} elevation that mediates the turning of growth cones.

Extracellular Ca^{2+} is required for the turning of growth cones of cultured *Xenopus* spinal neurons induced by several guidance

factors³⁻⁶, including netrin-1. Application of an extracellular gradient of netrin-1 at the growth cone triggers an elevation of cytosolic Ca^{2+} , which is required for both attractive and repulsive turning^{3,7}. Among various Ca^{2+} -permeant channels, TRP channels are good candidates for providing netrin-1-induced Ca^{2+} influx because of their Ca^{2+} permeability near the resting membrane potential^{1,2}. To monitor directly ion channel activities, we performed whole-cell patch recordings from the growth cones of cultured *Xenopus* spinal

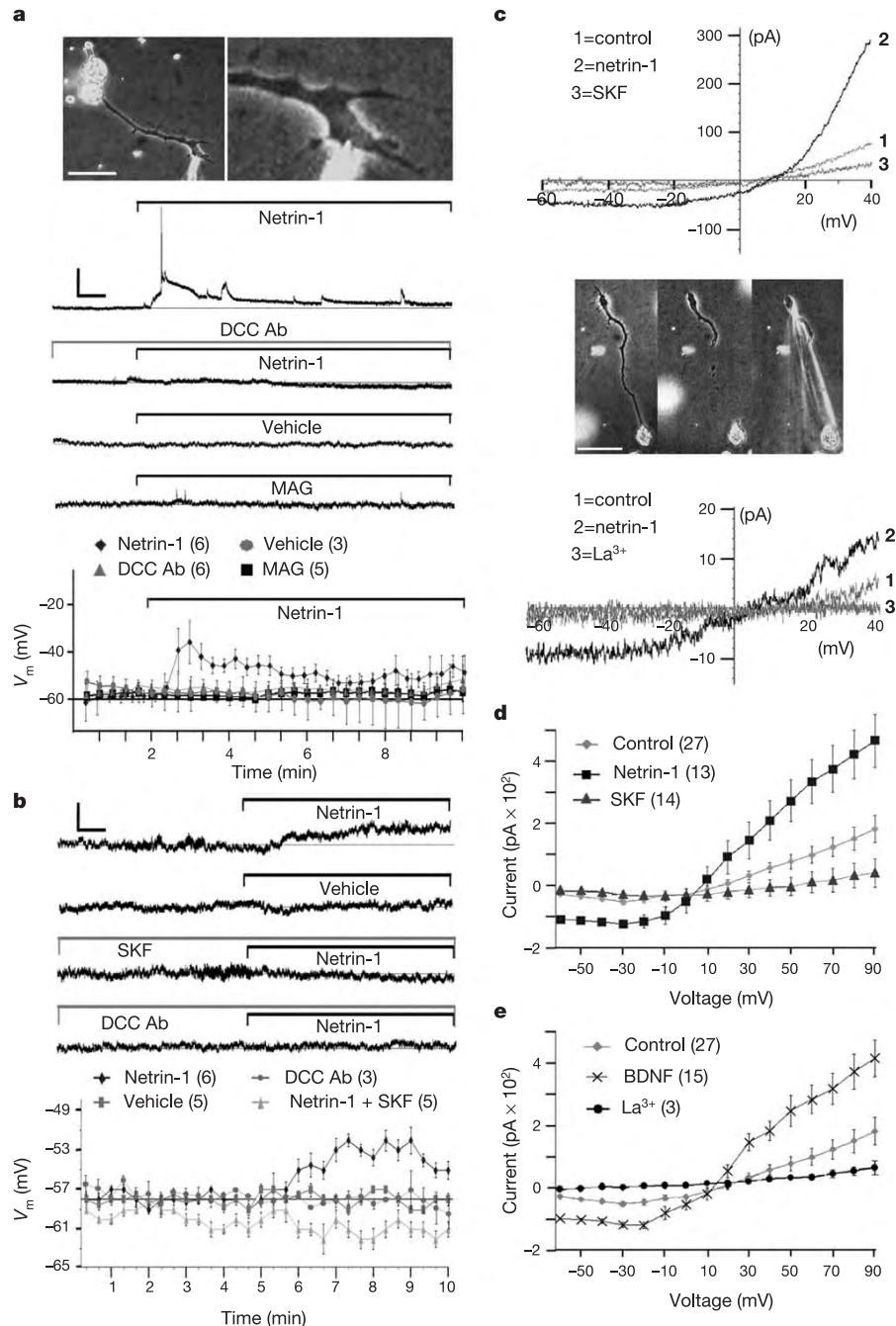


Figure 1 Membrane depolarization and putative TRPC currents induced by netrin-1 and BDNF. **a**, Image of a *Xenopus* neuron during whole-cell recording from the growth cone (top panel). Scale bar, 20 μm . The bottom panel shows examples of growth-cone membrane potential before and after application of netrin-1, vehicle or MAG, with or without anti-DCC antibody in the bath. The graph shows averaged membrane potential ($\pm$ s.e.m., over 20-ms bins). Scale bar: 40 mV, 1 min. **b**, Similar to the bottom panel of **a**, except that all major voltage-dependent channels were blocked. Scale bar: 10 mV, 1 min. **c**, I/V relationship of putative TRPC currents measured with a voltage ramp, before and

after netrin-1 application, and after subsequent perfusion with SKF-96365 (top panel). The middle panel shows images of a growth cone, before and after neurite transection and during recording. Scale, 20 μm . The bottom panel shows the I/V relationship of the putative TRPC current at a transected growth cone. **d**, **e**, Summary of putative growth-cone TRPC currents measured with step depolarizations, with or without netrin-1 (**d**) or BDNF (**e**), or with the guidance factor plus SKF-96365 (**d**) or La^{3+} (**e**). Error bars indicate s.e.m.

neurons (Fig. 1a, top panel), before and after the application of netrin-1. When a growth cone was patched and perforated to monitor the membrane potential (see Methods), application of netrin-1 caused a significant depolarization (Fig. 1a, bottom panel), which was absent when the culture was incubated with an antibody against DDC (deleted in colorectal cancer), the receptor for netrin-1. The effect of netrin-1 on membrane potential was specific, because no depolarization was induced by similar applications of the vehicle solution or the guidance factor myelin-associated glycoprotein (MAG; Fig. 1a, bottom panel). Under the condition where most voltage-dependent K^+ channels, Na^+ channels, and L-, N-, P- and T-type Ca^{2+} channels (VDCCs) were blocked by pharmacological agents (see Methods), there remained a slight but significant depolarization (3.6 ± 0.5 mV, $n = 5$), which was abolished by pre-incubation with SKF-96365 (25 μ M), a compound known to block TRP channels^{8–10} (Fig. 1b). This depolarization was due to specific netrin-1 action, because it was prevented by the presence of anti-DCC antibody. Although SKF-96365 is known to have nonspecific actions on other channels¹¹, these results raised the possibility that netrin-1, through its activation of DCC, triggers TRP current-like activity that depolarizes the growth cone.

To characterize further growth-cone channel activities, we directly examined TRP channels at the growth cone by isolating the putative TRP current with pharmacological agents that blocked the majority of voltage-dependent channels (see above). Shown in Fig. 1c (top panel) is an example of the current–voltage (I/V) relationship determined by using a voltage ramp. A small, putative TRP current was observed before exposure to netrin-1, and addition of netrin-1 in the bath (4 ng ml^{−1}) markedly elevated this current within 5 min. Further bath addition of SKF-96365 (25 μ M) reduced the current to a level below that found before the netrin-1 treatment. Because TRP channels are found in both the soma and growth cones of cultured hippocampal neurons¹², we performed recordings on growth cones severed from the soma. As shown in Fig. 1c (middle panel), the amplitude of the putative TRP current recorded in the isolated growth cone (together with a piece of neurite) was about tenfold smaller than that recorded from the growth cones of intact neurons (Fig. 1c, bottom panel), although the I/V relationship was similar. Thus, the TRP channel-like activity was present in the part of the neurite that includes the growth cone.

Quantitative I/V measurements by sequential step depolarizations are summarized in Fig. 1d for all recordings made before and after netrin-1 application, and after further addition of SKF-96365. The effect of netrin-1 on putative TRP currents in these neurons is not unique, because exposure to brain-derived neurotrophic factor (BDNF), another chemoattractant for these growth cones¹³, also led to a similar enhancement of these currents (Fig. 1e). The effect of BDNF was largely abolished in the presence of SKF-96365 (data not shown) and La^{3+} (50 μ M), another agent generally used to block TRP currents^{14,15} (Fig. 1e). This effect of BDNF is consistent with the finding that BDNF–TrkB signalling leads to TRP channel activation in rat pontine neurons⁹. Taken together, our results support the notion that TRP channels are present at the growth cone and their activity can be elevated by netrin-1 and BDNF. There is evidence for the involvement of non-voltage-gated, Ca^{2+} -permeant channels in substrate-induced repulsive growth-cone turning¹⁶. Recent studies of BDNF guidance of the growth cone of cerebellar granule cells in culture have also shown that TRPC channels are required¹⁷.

To determine whether TRP channels are involved in growth-cone guidance in *Xenopus* spinal neurons, we assayed the growth-cone turning induced by netrin-1 and BDNF. As shown in Fig. 2a, c, a gradient of netrin-1 or BDNF near the growth cone induced attractive turning^{3,4,13}, whereas bath addition of SKF-96365 (15 μ M) abolished the turning (Fig. 2b, d) without significantly affecting the rate of neurite extension (Fig. 2e). Thus, TRP channels seem to be involved in growth-cone guidance by netrin-1 and BDNF.

Only one member of the TRP family has been identified in

Xenopus: a TRPC1 homologue *Xenopus* TRP-1 (refs 18, 19), which we refer to hereafter as xTRPC1. In mammalian neurons, TRPC1 forms functional channels as homomers or as heteromeric complexes with other TRPC subunits^{8,12,20–27}. The I/V properties of putative TRP currents that we found at the growth cone resemble those reported for heteromeric TRPC channels^{20–23,25–27}, suggesting the existence of other *Xenopus* TRPC proteins. To examine whether xTRPC1 is necessary for the putative TRP current observed, a morpholino oligonucleotide (MO) was designed to inhibit the translation of xTRPC1 messenger RNA. Injection of the MO into freshly fertilized *Xenopus* eggs caused a significant knockdown in xTRPC1 expression during early development, as shown by western blotting of total protein extracts from 1.5 d embryos or isolated spinal cords (Fig. 3a; see also Supplementary Fig. 1). The level of xTRPC1 expression in MO-injected embryos, normalized to that of β -tubulin, was $32 \pm 7\%$ of that found in control (non-injected) embryos of the same developmental stage. In contrast, no down-

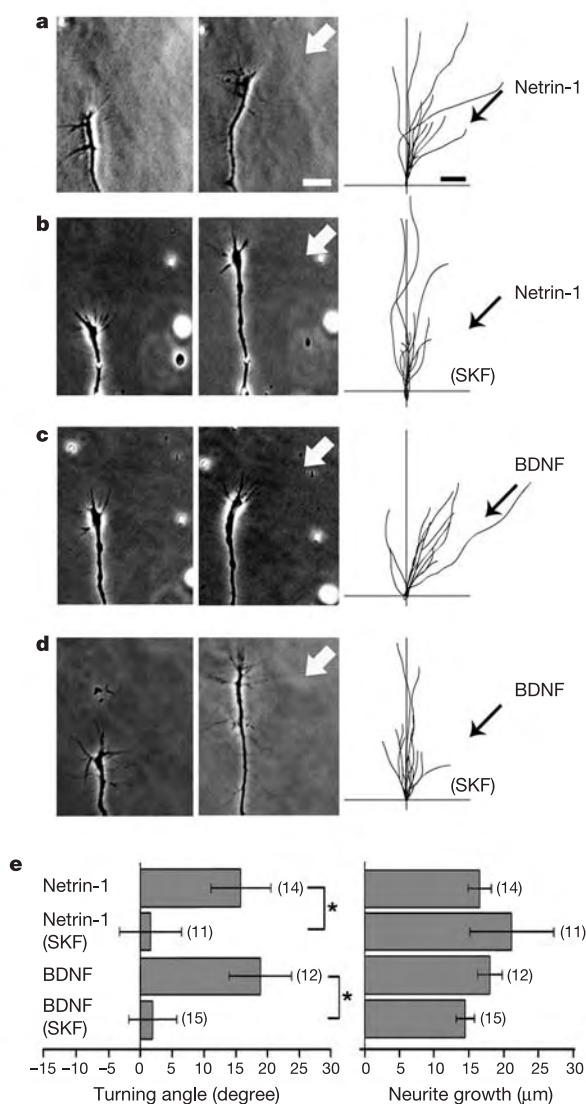


Figure 2 Growth-cone turning induced by netrin-1 or BDNF, and the effect of SKF-96365. **a–d**, Images of *Xenopus* neurons before and 1-h after exposure to a gradient of netrin-1 (**a, b**) or BDNF (**c, d**), without (**a, c**) or with (**b, d**) SKF-96365 in the bath. The arrow indicates the direction of the gradient. Scale bar, 20 μ m. Traces depict the growth-cone trajectory during the turning assay (scale, 10 μ m). **e**, Summary of all data on the average turning angle and the neurite extension, under conditions described in **a–d**. Error bars indicate s.e.m. (asterisk, $P < 0.03$; Mann–Whitney U -test).

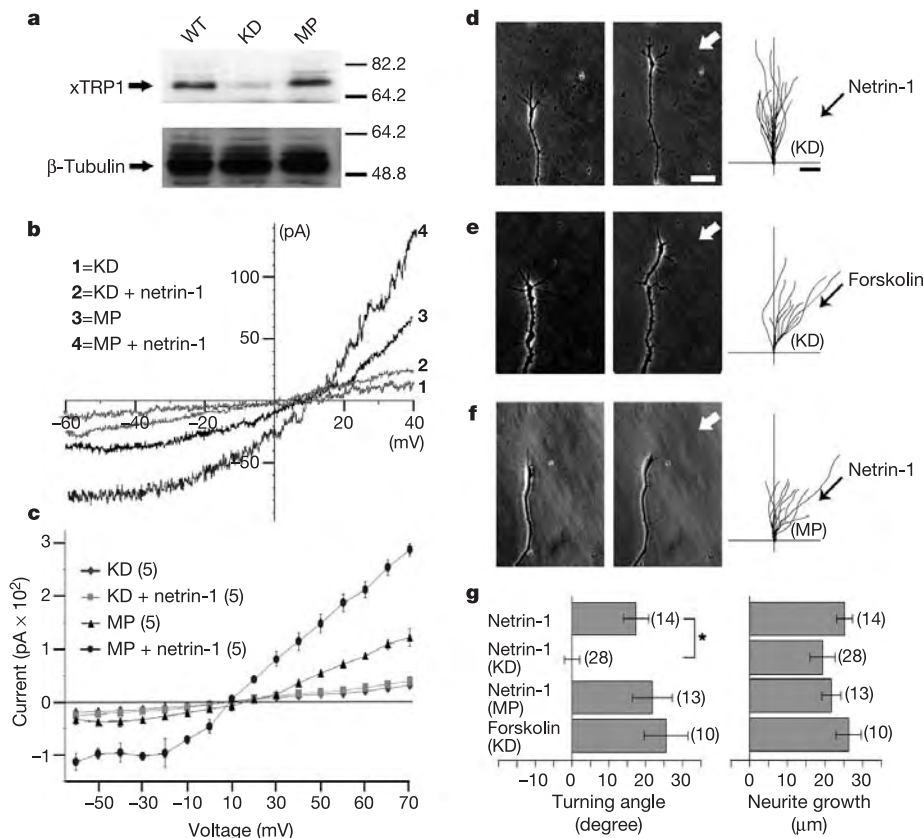


Figure 3 Effects of xTRPC1 MO on TRPC currents and growth-cone turning induced by netrin-1. **a**, Western blots of xTRPC1 from total protein extracts of dissected *Xenopus* neural tubes. WT, non-injected (wild type); KD, xTRPC1 MO-injected; MP, control misprimed MO-injected. The bottom panel shows β -tubulin loading control. **b**, TRPC currents measured with a voltage ramp in KD and MP growth cones, with or without netrin-1 application. **c**, Summary of TRPC currents (with step depolarizations) from

experiments as in **b**. **d-f**, Growth-cone turning assay for KD (**d, e**) and MD (**f**) neurons in a gradient of netrin-1 (**d, f**) or forskolin (**e**), depicted as in Fig. 2. **g**, Average turning angle and neurite extension for conditions described in **d-f**, and for control (non-fluorescent) neurons in the same culture as for KD and MP. Error bars indicate s.e.m. (asterisk, $P < 0.0002$; Mann-Whitney *U*-test).

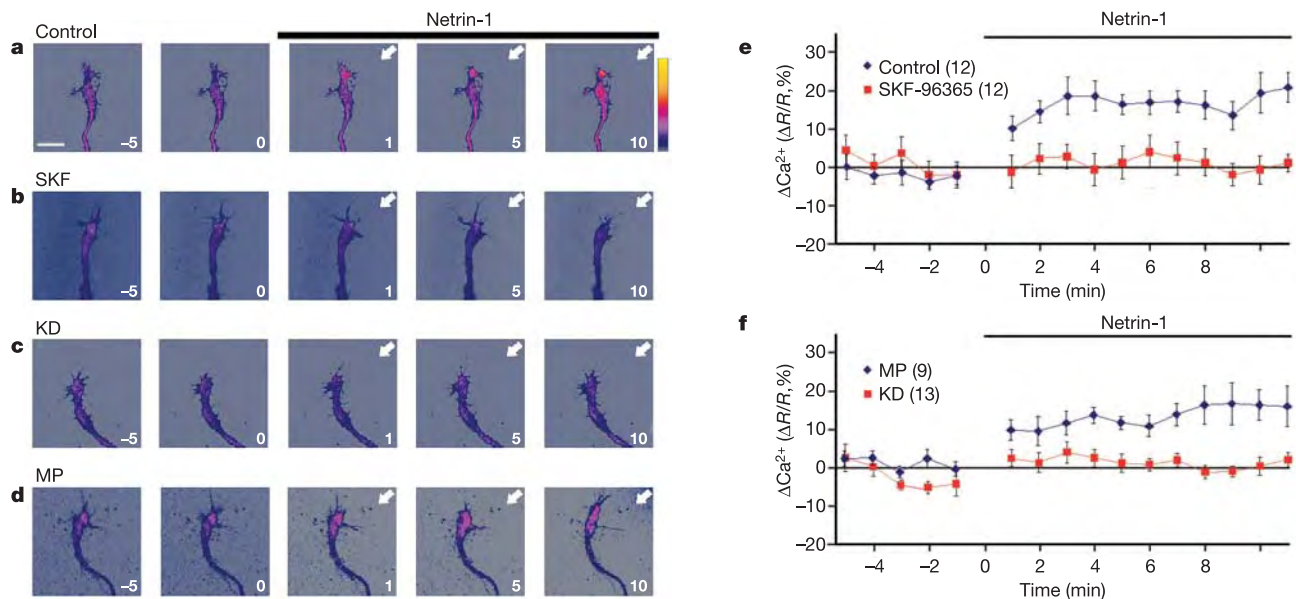


Figure 4 xTRPC1 mediates netrin-1-induced Ca^{2+} elevation in the growth cone. **a-d**, *Xenopus* growth cones were co-injected with Oregon green-BAPTA and Fura red (see Methods). Images depict the ratio of green and red fluorescence at different times (in minutes) before and after netrin-1 application, coded in linear scale by pseudo-colours (colour bar). Arrows indicate the direction of the netrin-1 gradient. Examples shown are

neurons from a non-injected embryo without (Control) or with SKF-96365 in the bath (SKF), and from a xTRPC1 MO-injected (KD) or misprimed MO-injected (MP) embryo. Scale bar, 10 μ m. **e, f**, Summary of netrin-1-induced Ca^{2+} changes at the growth cone as in **a-d**. Percentage changes of the fluorescence ratio ($\Delta R/R$) were calculated for each growth cone before averaging. Error bars indicate s.e.m.

regulation was observed in embryos injected with a MO of a similar sequence but designed with five misprimed nucleotides (Fig. 3a). Neurons from embryos injected with the xTRPC1 knockdown MO showed normal morphology and a small but significant increase in neurite growth (Supplementary Fig. 2); however, putative TRPC currents measured at the growth cones of these neurons were significantly lower than those found in neurons from embryos loaded with the misprimed MO (Fig. 3b, c). The residual current did not exhibit rectification at higher membrane voltages, as normally observed for the putative TRPC currents in control neurons (Fig. 3b). We note that the amount of MO in cultured neurons, as detected by MO fluorescence, was highly variable. We chose to examine neurons with higher levels of fluorescence for all our functional assays. The percentage reduction of xTRPC1 in these cells is likely to be higher than that shown by the western blot.

To assess the functional consequence of the xTRPC1 knockdown, we examined the turning behaviour of growth cones of xTRPC1 MO-loaded neurons. We found that the growth cones of xTRPC1 MO-loaded neurons failed to respond to the netrin-1 gradient

(Fig. 3d), similar to growth cones of neurons incubated with SKF-96365. Moreover, the effect of xTRPC1 MO was specific, because growth cones of neurons loaded with misprimed MO exhibited normal netrin-1-induced turning (Fig. 3f). The xTRPC1 MO did not disrupt the cellular mechanism required for the turning response, because xTRPC1 knockdown neurons showed attractive turning in a forskolin gradient (Fig. 3e; see also ref. 28). Taken together, these results indicate that xTRPC1 is essential for the attractive growth-cone turning induced by netrin-1.

Although Ca^{2+} influx was shown to be necessary for growth-cone turning induced by netrin-1, only L-type VDCCs have been implicated^{3,6}. Pharmacological blockade of L-type VDCCs converts netrin-1-induced attraction to repulsion³, but the repulsion still requires the presence of extracellular Ca^{2+} , suggesting that other Ca^{2+} -permeant channels are involved. Our present results suggest that TRP channels are the likely candidates for mediating this residual Ca^{2+} influx required for netrin-1-induced repulsion. Fluorescence ratio imaging of Ca^{2+} showed that blocking L-type VDCCs only partially reduced netrin-1-induced Ca^{2+} elevation in

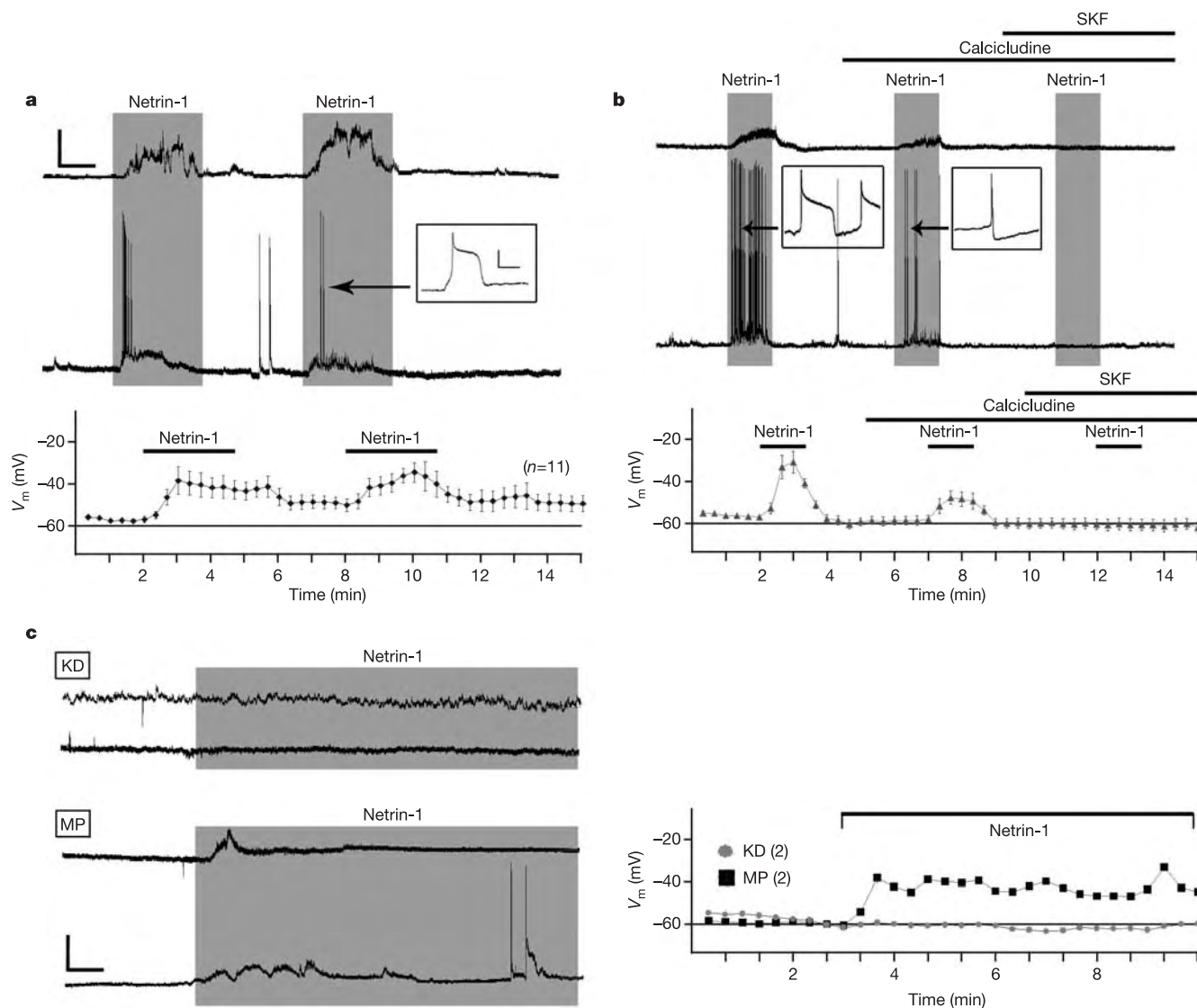


Figure 5 xTRPC1 mediates netrin-1-induced membrane depolarization. **a**, Two examples of membrane potential changes recorded from the growth cone before and after local netrin-1 application (for 2.5 min). Scale bar: 40 mV, 1 min. The inset shows a sample of an action potential. Scale bar: 50 mV, 200 ms. The bottom image shows an average from all cells recorded ($\pm$ s.e.m.). **b**, Similar to **a** except that netrin-1 application was 1 min in duration, and calcicludine (which blocks L-, N- and P-type VDCCs) and SKF-96365 were

applied to the bath before the second and the third netrin-1 application, respectively. The inset shows action potentials before and after applying calcicludine. **c**, Similar to **a** except that netrin-1 was applied for 7 min to neurons from embryos injected with xTRPC1 MO (KD) or misprimed MO (MP). The summary graph shows an average of two recordings for each condition. Scale bar: 40 mV, 0.5 min.

Xenopus growth cones³. In the present study, we found a significant Ca^{2+} elevation in growth cones within 1 min of netrin-1 application (Fig. 4a, e), but this Ca^{2+} elevation was essentially absent in the presence of SKF-96365 (Fig. 4b, e) or in neurons loaded with the xTRPC1 MO (Fig. 4c, f), but not in neurons loaded with the misprimed MO. Thus, xTRPC1 is essential for netrin-1-induced Ca^{2+} elevation in these *Xenopus* neurons. Furthermore, together with previous data^{3,6}, our findings suggest that xTRPC1 channels are probably responsible for allowing Ca^{2+} influx through not only themselves but also VDCCs.

Recent studies have shown that step depolarizations applied to the growth cones of *Xenopus* spinal neurons result in larger L-type VDCC currents after exposure to netrin-1 (ref. 6), but how these VDCCs are activated by netrin-1 under normal physiological conditions (in the absence of artificial depolarization) remains unclear. As shown earlier (Fig. 1b), netrin-1 activation of TRPC-like currents induces a membrane depolarization near the resting potential, which may, in turn, activate Na^+ channels and other low-threshold VDCCs²⁹, leading to further membrane depolarizations and activation of high-threshold (L-type) VDCCs. As shown in Fig. 5a, substantial depolarization of about 20 mV can be repeatedly induced by netrin-1, resulting in the initiation of action potentials in about 15% (5 out of 38) of growth cones recorded. Even in cells without action potential initiation, the average netrin-1-induced depolarization was close to 20 mV, which depolarizes the cell from a resting potential of -55.7 ± 0.6 mV ($n = 46$) to a range sufficient to activate L-type VDCCs (see ref. 6). The action potentials induced by netrin-1 exhibited a plateau depolarization characteristic of L-type VDCC activity, which disappeared when the L-type VDCC blocker calcicludine was added into the bath (Fig. 5b). Further addition of SKF-96365 abolished netrin-1-induced depolarizations as well as action potentials, suggesting that the TRPC current provides the initial trigger for substantial membrane depolarization, including action potentials. That TRPC channels are essential for triggering netrin-1-induced depolarization is further supported by the finding that in physiological levels of saline, netrin-1 induced no significant depolarization in neurons loaded with xTRPC1 MO, but substantial depolarization in neurons loaded with misprimed MO (Fig. 5c). Furthermore, SKF-96365 by itself completely blocked the netrin-1-induced depolarization and did not affect voltage-dependent channels in the growth cone (Supplementary Fig. 3).

The above results show that, besides their own contribution to Ca^{2+} influx, TRP channels may further augment Ca^{2+} influx through other VDCCs, including L-type VDCCs, by triggering a substantial membrane depolarization. Pharmacological blockade or downregulation of TRP channels resulted in complete elimination of netrin-1-induced Ca^{2+} elevation in growth cones (Fig. 4), but blocking L-type VDCCs only partially reduced it³, consistent with the idea that TRP currents are responsible for activating L-type VDCCs. Thus, TRP channel activity serves as the early trigger of Ca^{2+} elevation that mediates netrin-1-induced growth-cone turning. Our results illustrate the potential role of TRPC channels in transducing extracellular signals into cytoplasmic activities underlying directed cell motility. The expanding functions of TRP channels in sensory processes now include the chemosensory response of the nerve growth cone to extracellular guidance factors. □

Methods

Culture preparations

Cultures of spinal neurons were prepared from neural tube tissues of stage 22–25 *Xenopus* embryos, as described previously^{28,29}. Cultures were used 12–26 h after plating at room temperature (20–22 °C). Culture medium consisted of 86% (v/v) L-15 Leibovitz medium (GIBCO/Sigma), 0.4% (v/v) fetal bovine serum (Gemini Bioproducts), 0.2% penicillin/streptomycin (GIBCO/Sigma) and 12% (v/v) CM-2. The composition of CM-2 was (in mM): 0.44 NaH_2PO_4 , 1.34 Na_2HPO_4 , 1.26 CaCl_2 , 10 glucose, 5 sodium pyruvate and 4 HEPES (at pH 7.4).

Electrophysiology

Whole-cell recordings using amphotericin B perforation ($200 \mu\text{g ml}^{-1}$) were made on the growth cone using a patch-clamp amplifier (Axopatch 200B, Axon Inst.). Recording pipettes had resistances of 5–8 M Ω and were filled with internal solution consisting of (in mM) 125 K^+ gluconate, 20 KCl, 10 HEPES, 0.2 EGTA, 1 MgCl_2 , 1 NaCl (pH 7.4, osmolality 290–296 mosM). The bath solution contained (in mM): 140 NaCl, 35 TEA-Cl, 1 CaCl_2 , 10 HEPES. To isolate the putative TRP current, K^+ gluconate and KCl were replaced with CsCl in the internal solution, and the following blockers were added in the bath: tetrodotoxin (1 μM ; Na channels, Sigma), calcicludine (~ 200 nM; L-, N- and P-type VDCCs, Calbiochem)³⁰, Ni^{2+} (50 μM ; T-type VDCCs, Sigma), mibefradil (5 μM ; T-type VDCCs, Sigma) and pimoide (70 μM ; T-type VDCCs, Sigma). Anti-DCC antibody (Calbiochem) was used to block DCC receptors. For I/V measurement, step depolarizations from -60 to $+90$ mV (10 mV steps, 100 ms) or a voltage ramp from -60 to $+40$ mV (100 ms) were generated using P-Clamp 8.0 (Axon Instr.). Data were collected at 10 kHz and filtered at 2 kHz. Netrin-1 and BDNF were applied in the bath (final concentration of 4 and 20 ng ml^{-1} , respectively) for recordings from growth cones, either by bath application or by application from a micropipette similar to the turning assay.

Morpholino knockdown and western blotting

Morpholino oligonucleotide (MO) was designed to target a 25-nucleotide sequence in the 5' untranslated region of the reported gene sequence of *Xenopus* TRP-1 (refs 18, 19) (5'-CTCTGATAAAGAGCAGCCATGATGA-3'). A control MO was based on the previous morpholino sequence, with five randomized (misprimed) nucleotides (5'-CTGTCATAAAGACAGGCATCATGA-3'). These MOs were designed by and purchased from Genetools and were conjugated to lissamine for fluorescence visualization. Ten nanolitres of 2 mM stock MO solution were injected into 1- or 2-cell stage *Xenopus* embryos. Injected embryos were dissected at stage 22–25 and spinal neuron cultures were prepared. Lissamine fluorescence was used to identify MO-loaded cells before the experiment was conducted. For western blotting, total protein extracts were made from ~ 50 embryos (stage 25–28) each for the non-injected, anti-xTRPC1 MO-injected and control MO-injected group. In some experiments, protein extracts were prepared from 100 spinal cords from each group (stage 25–28), which were isolated by using 0.25% dispase (Calbiochem) and 0.25% collagenase (Calbiochem). The samples were quantified by BCA protein assay (Pierce) and a standardized amount (50 μg) was loaded on pre-cast polyacrylamide gels (BioRAD). Samples were run at 200 mV and a semi-dry transfer system was used to blot proteins onto nitrocellulose membranes (BioRAD); they were then probed with xTRPC1 antibody (gift from G. J. Barritt), visualized by ECL plus western blotting detection (Amersham Biosciences), then stripped and re-probed with β -tubulin antibody (Santa Cruz Biotechnology).

Growth-cone turning assay

Microscopic gradients of guidance factors were produced as previously described^{3,28}. The concentration of netrin-1 and BDNF in the micropipette was 4 and 20 $\mu\text{g ml}^{-1}$, respectively, resulting in an average extracellular concentration at the growth cone of about 4 and 20 ng ml^{-1} (refs 3, 28). Images of neurites were recorded and analysed using Scion Image 4.0.2. The turning angle was defined by the angle between the original direction of neurite extension and a straight line connecting the growth-cone positions at the onset and the end of the 1-h period. Only those growth cones with net extension $> 7 \mu\text{m}$ over the 1-h period were analysed.

Calcium imaging

For cultured *Xenopus* spinal neurons, cells were microinjected with Fura red and Oregon green 488 BAPTA-1 conjugated to dextran (Molecular Probes). Calcium imaging was done using a Leica confocal microscope (LCS SP) as described previously^{3,4}. Excitation was at 488 nm, and Oregon green and Fura red emission signals were collected at 500–560 nm and 605–700 nm, respectively; fluorescence images (128 $\times$ 128 pixels; voxel size 207 nm) were collected every 30 s and analysed using NIH Image J software (<http://rsb.info.nih.gov/ij/>). The mean fluorescence intensity was measured over a fixed square area that covered the entire growth cone. Ratios of the normalized Oregon green and Fura red fluorescence were used to control for fluctuations in growth-cone thickness or changes in focal plane.

Received 19 December 2004; accepted 21 February 2005; doi:10.1038/nature03478.

Published online 9 March 2005.

- Clapham, D. E. TRP channels as cellular sensors. *Nature* **426**, 517–524 (2003).
- Montell, C., Birnbaumer, L. & Flockerzi, V. The TRP channels, a remarkably functional family. *Cell* **108**, 595–598 (2002).
- Hong, K., Nishiyama, M., Henley, J., Tessier-Lavigne, M. & Poo, M. Calcium signalling in the guidance of nerve growth by netrin-1. *Nature* **403**, 93–98 (2000).
- Henley, J. & Poo, M. M. Calcium mediates bidirectional growth cone turning induced by myelin-associated glycoprotein. *Neuron* **44**, 909–916 (2004).
- Henley, J. & Poo, M. M. Guiding neuronal growth cones using Ca^{2+} signals. *Trends Cell Biol.* **14**, 320–330 (2004).
- Nishiyama, M. *et al.* Cyclic AMP/GMP-dependent modulation of Ca^{2+} channels sets the polarity of nerve growth-cone turning. *Nature* **423**, 990–995 (2003).
- Ming, G. L. *et al.* cAMP-dependent growth cone guidance by netrin-1. *Neuron* **19**, 1225–1235 (1997).
- Kim, S. J. *et al.* Activation of the TRPC1 cation channel by metabotropic glutamate receptor mGluR1. *Nature* **426**, 285–291 (2003).
- Li, H. S., Xu, X. Z. & Montell, C. Activation of a TRPC3-dependent cation current through the neurotrophin BDNF. *Neuron* **24**, 261–273 (1999).
- Zhu, X., Jiang, M. & Birnbaumer, L. Receptor-activated Ca^{2+} influx via human Trp3 stably expressed in human embryonic kidney (HEK)293 cells. Evidence for a non-capacitative Ca^{2+} entry. *J. Biol. Chem.* **273**, 133–142 (1998).

11. Schwarz, G., Droogmans, G. & Nilius, B. Multiple effects of SKF 96365 on ionic currents and intracellular calcium in human endothelial cells. *Cell Calcium* **15**, 45–54 (1994).
12. Greka, A., Navarro, B., Oancea, E., Duggan, A. & Clapham, D. E. TRPC5 is a regulator of hippocampal neurite length and growth cone morphology. *Nature Neurosci.* **6**, 837–845 (2003).
13. Song, H. J., Ming, G. L. & Poo, M. M. cAMP-induced switching in turning direction of nerve growth cones. *Nature* **388**, 275–279 (1997).
14. Jung, S. *et al.* Lanthanides potentiate TRPC5 currents by an action at extracellular sites close to the pore mouth. *J. Biol. Chem.* **278**, 3562–3571 (2003).
15. Inoue, R. *et al.* The transient receptor potential protein homologue TRP6 is the essential component of vascular $\alpha(1)$ -adrenoceptor-activated Ca^{2+} -permeable cation channel. *Circ. Res.* **88**, 325–332 (2001).
16. Gomez, T. M., Robles, E., Poo, M. & Spitzer, N. C. Filopodial calcium transients promote substrate-dependent growth cone turning. *Science* **291**, 1983–1987 (2001).
17. Li, Y. *et al.* Essential role of TRPC channels in the guidance of nerve growth cones by brain-derived neurotrophic factor. *Nature* doi:10.1038/nature03477 (this issue).
18. Bobanovic, L. K. *et al.* Molecular cloning and immunolocalization of a novel vertebrate trp homologue from *Xenopus*. *Biochem. J.* **340**, 593–599 (1999).
19. Brereton, H. M., Harland, M. L., Auld, A. M. & Barritt, G. J. Evidence that the TRP-1 protein is unlikely to account for store-operated Ca^{2+} inflow in *Xenopus laevis* oocytes. *Mol. Cell. Biochem.* **214**, 63–74 (2000).
20. Lintschinger, B. *et al.* Coassembly of Trp1 and Trp3 proteins generates diacylglycerol- and Ca^{2+} -sensitive cation channels. *J. Biol. Chem.* **275**, 27799–27805 (2000).
21. Strubing, C., Krapivinsky, G., Krapivinsky, L. & Clapham, D. E. TRPC1 and TRPC5 form a novel cation channel in mammalian brain. *Neuron* **29**, 645–655 (2001).
22. Strubing, C., Krapivinsky, G., Krapivinsky, L. & Clapham, D. E. Formation of novel TRPC channels by complex subunit interactions in embryonic brain. *J. Biol. Chem.* **278**, 39014–39019 (2003).
23. Xu, X. Z., Li, H. S., Guggino, W. B. & Montell, C. Coassembly of TRP and TRPL produces a distinct store-operated conductance. *Cell* **89**, 1155–1164 (1997).
24. Amiri, H., Schultz, G. & Schaefer, M. FRET-based analysis of TRPC subunit stoichiometry. *Cell Calcium* **33**, 463–470 (2003).
25. Schaefer, M., Plant, T. D., Stresow, N., Albrecht, N. & Schultz, G. Functional differences between TRPC4 splice variants. *J. Biol. Chem.* **277**, 3752–3759 (2002).
26. Goel, M., Sinkins, W. G. & Schilling, W. P. Selective association of TRPC channel subunits in rat brain synaptosomes. *J. Biol. Chem.* **277**, 48303–48310 (2002).
27. Hofmann, T., Schaefer, M., Schultz, G. & Gudermann, T. Subunit composition of mammalian transient receptor potential channels in living cells. *Proc. Natl Acad. Sci. USA* **99**, 7461–7466 (2002).
28. Lohof, A. M. *et al.* Asymmetric modulation of cytosolic cAMP activity induces growth cone turning. *J. Neurosci.* **12**, 1253–1261 (1992).
29. Gu, X. & Spitzer, N. C. Low-threshold Ca^{2+} current and its role in spontaneous elevations of intracellular Ca^{2+} in developing *Xenopus* neurons. *J. Neurosci.* **13**, 4936–4948 (1993).
30. Schweitz, H. *et al.* Calcicludine, a venom peptide of the Kunitz-type protease inhibitor family, is a potent blocker of high-threshold Ca^{2+} channels with a high affinity for L-type channels in cerebellar granule neurons. *Proc. Natl Acad. Sci. USA* **91**, 878–882 (1994).

Supplementary Information accompanies the paper on www.nature.com/nature.

Acknowledgements We thank G. J. Barritt for the gift of xTRP-1 antibody. This work was supported by a grant from NIH and a NSF predoctoral fellowship (to G.X.W.).

Competing interests statement The authors declare that they have no competing financial interests.

Correspondence and requests for materials should be addressed to M.P. (mpoo@berkeley.edu).

Sheep retrovirus structural protein induces lung tumours

Sarah K. Wootton*, Christine L. Halbert* & A. Dusty Miller

Fred Hutchinson Cancer Research Center, 1100 Fairview Avenue North, Seattle, Washington 98109, USA

* These authors contributed equally to this work

Jaagsiekte sheep retrovirus (JSRV) causes a contagious lung cancer in sheep and goats, with significant animal health and economic consequences¹. The host range of JSRV is in part limited by species-specific differences in the virus entry receptor, hyaluronidase 2 (Hyal2), which is not functional as a receptor in mice but is functional in humans². Sheep are immunotolerant of JSRV because of the expression of closely related endogenous retroviruses^{3,4}, which are not present in humans and most other species, and this may facilitate oncogenesis. Here we show that

expression of the JSRV envelope (Env) protein alone in lungs of mice, by using a replication-incompetent adeno-associated virus vector, results in tumours with a bronchiolo-alveolar localization like those seen in sheep. Whereas lethal disease was observed in immunodeficient mice, tumour development was almost entirely blocked in immunocompetent mice. Our results provide a rare example of an oncogenic viral structural protein, show that interaction of the viral Env protein with the virus entry receptor Hyal2 is not required for tumorigenesis, and indicate that immune recognition of Env can protect against JSRV tumorigenesis.

Oncogenic retroviruses are known to cause cancer by the acquisition and expression of host-derived oncogenes, by the insertional activation of host cell oncogenes or by the expression of auxiliary viral oncogenes such as the *tax* gene of human T-cell leukaemia virus. JSRV is a simple retrovirus (Fig. 1) that does not express a host-derived or auxiliary oncogene and can induce lung tumours in as little as 10 days⁵, a much shorter latency than typically found for the insertional activation of host oncogenes by other retroviruses. The mechanism of oncogenesis is unknown, but the JSRV Env protein has been found to transform cells in culture^{2,6–8}. One mechanism of transformation involves activation of the phosphoinositide-3-OH kinase (PI3K)/Akt pathway and is dependent on the presence of the cytoplasmic tail of Env^{8–10}, and the other involves Env binding to Hyal2, Hyal2 degradation, and activation of the RON receptor tyrosine kinase, which is normally suppressed by Hyal2 (ref. 11).

Further studies of Env oncogenesis in animals are limited by the difficulty and expense of experimentation with a contagious oncogenic virus in sheep and by the inability of JSRV to infect convenient rodent animal models such as mice. However, we have found that adeno-associated virus (AAV) vectors made with AAV type 6 capsid proteins (AAV6 vectors) can promote long-term gene expression in all epithelial cell types in mouse lung¹². To test whether Env alone would induce lung tumours we administered a mixture of 5×10^{10} vector genomes of an AAV6 vector that expressed Env (ARJenv; Fig. 1) and 5×10^9 vector genomes of an AAV6 vector that expressed human placental alkaline phosphatase (AP) (ARAP4; Fig. 1) (ref. 13) to the noses of lightly anaesthetized mice and monitored the mice for AP expression and tumour development. The ARAP4 vector was included to confirm that vector transduction had occurred. We used 8-week-old immunodeficient (Rag2-knock-out) C57BL/6 mice as recipients to avoid an immune response that might eliminate Env-expressing cells and because C57BL/6 mice are resistant to the development of spontaneous lung cancer¹⁴. Individual mice were killed at 2, 2.5, 5 and 6 months after vector exposure and their lungs were fixed and stained for AP expression. Lung tumours were present in all mice and increased in size and number with time (2 months, Fig. 2a, e; 6 months, Fig. 2b, f;

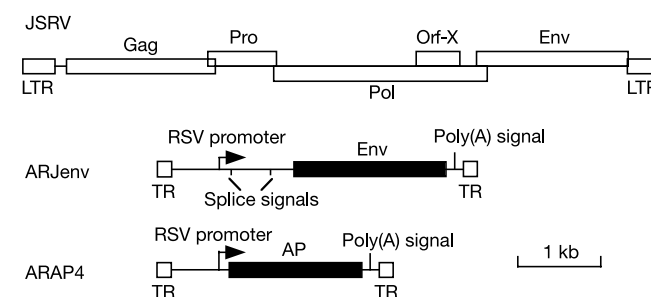


Figure 1 Scale drawings of the JSRV genome and the AAV vectors encoding JSRV Env (ARJenv) and AP (ARAP4). The JSRV-coding regions are staggered vertically to indicate the three different reading frames that encode the proteins. Gag, core polyprotein; kb, kilobase; LTR, retroviral long terminal repeat; Orf-X, open reading frame of unknown function; Pol, polymerase; Poly(A) signal, polyadenylation signal; Pro, protease and dUTPase; RSV, Rous sarcoma virus; TR, AAV terminal repeat.

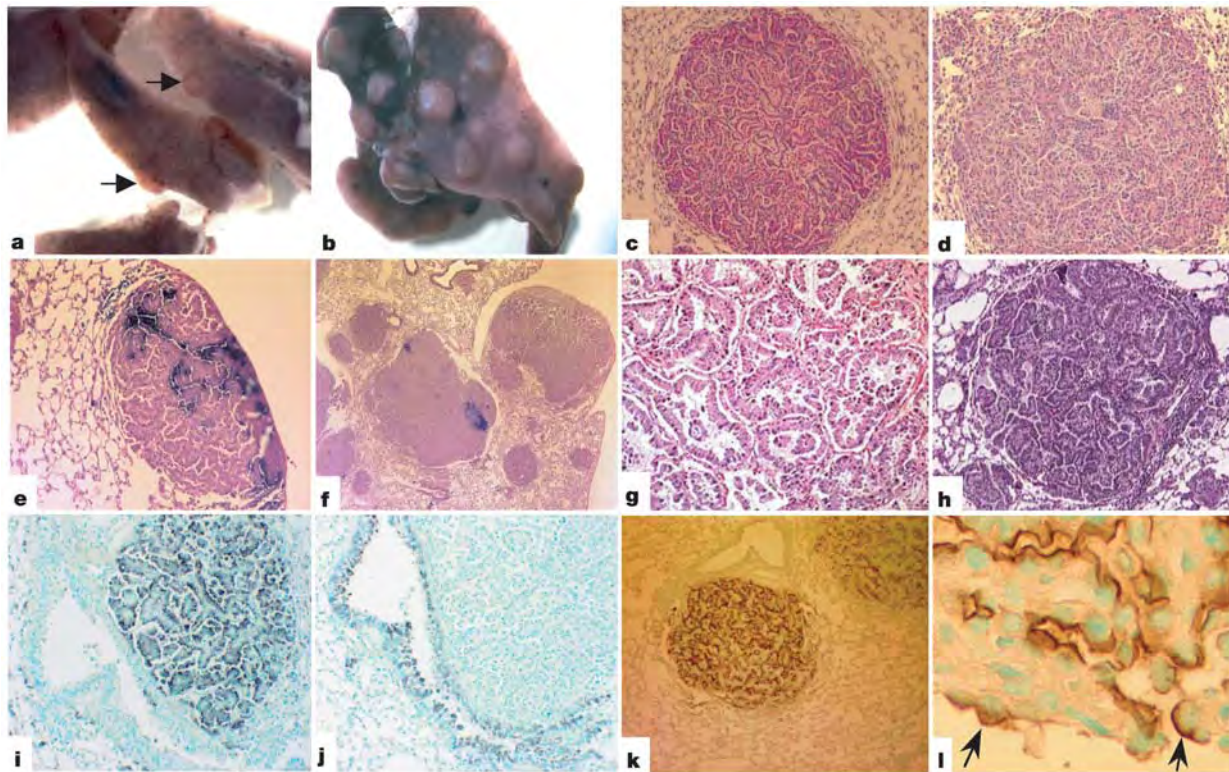


Figure 2 Characteristics of lung tumours induced by JSRV Env in mice. **a**, Lung tumours (arrows) 2 months after vector exposure. **b**, Lung tumours at 6 months. **c**, Papillary adenoma surrounded by normal peripheral lung tissue (compressed near tumour) at 2.5 months. **d**, Adenocarcinoma at 6 months. **e**, Surface tumour at 2 months. **f**, Multiple tumours at 6 months. **g**, Human lung peripheral adenocarcinoma. **h**, Lung tumour in a lamb 2 months after JSRV infection. **i**, SP-C immunostaining. Tumour and type II cells in

alveoli are stained but nearby airway is not. **j**, CC10 immunostaining of the tumour shown in **i**. Tumour is not stained but Clara cells lining the nearby airway are. **k**, JSRV Env immunostaining. **l**, Higher-magnification view of apical membrane Env staining. Arrows point to Env staining of apical membranes that face away from lung tissue and towards airways. Staining in **c–h** was with haematoxylin and eosin.

Table 1). AP staining was visible in some tumours (dark blue/black stain in Fig. 2e, f) and a few tumours stained uniformly for AP (not shown), showing that occasional tumour progenitor cells were transduced by both vectors. The animal killed at 6 months was severely underweight (21 g versus 35 g each for two age-matched mice that received control AAV6 vectors) and was experiencing breathing difficulties and signs of distress that necessitated euthanasia. No tumour or evidence of disease was seen in animals treated identically apart from receiving an AAV6 vector (ARJenvF) that expressed a transformation-defective JSRV Env instead of the vector encoding the active Env (fewer than two tumours per cm² in histological sections of lungs of individual animals killed at 2, 2.5, 5 and 6 months), revealing a highly significant difference in tumour number between mice receiving Env and those receiving the control vector ($P = 0.01$; two-tailed t -test).

Tumour histology revealed papillary adenomas (Fig. 2c) progressing to adenocarcinoma (Fig. 2d) at later time points. The mouse tumour histology resembled human peripheral adenocarcinoma (Fig. 2g) and that of tumours arising in sheep after experimental infection with JSRV (Fig. 2h). All mouse tumours expressed surfactant protein C (SP-C) (Fig. 2i), a marker for alveolar type II pneumocytes. With the exception of occasional small areas of staining, tumours did not express Clara cell 10-kDa (CC10) antigen, a marker for bronchiolar non-ciliated Clara cells (Fig. 2j). This pattern resembles that seen for JSRV-induced tumours in sheep¹⁵ and indicates that the primary target for Env oncogenesis might be the type II pneumocyte, the progenitor of type I pneumocytes that line the alveoli. Tumours were not seen in large airways or in the noses of these mice, even though the ARAP4 vector was able to transduce these cells as measured by the expression of AP (ref. 12

and data not shown), and by the expression of spliced Env RNA of the expected size in these tissues (Fig. 3).

We next tested whether immune function might influence Env tumorigenesis in mice. We exposed normal 8-week-old C57BL/6 mice to a mixture of 5×10^{10} vector genomes of the ARJenv vector and 5×10^9 vector genomes of the ARAP4 vector, and killed two mice each at 2 and 4 months after exposure. None of these animals showed signs of disease. A few small tumours were found in one animal killed at 2 months, whereas the rest of the animals were tumour-free (Table 1). Staining of the lungs for AP revealed similar transduction by the ARAP4 vector to that in the identically treated immunodeficient mice described above. We tested the mice for serum antibodies against JSRV Env by measuring the ability of the serum to neutralize the infectivity of a retroviral vector bearing JSRV Env [LAPSN(PJ4)] (ref. 16). Serum harvested at 2.5 and 4

Table 1 Tumour number and area in lungs of mice receiving the ARJenv vector

Mouse strain	Time (months)	Tumours/cm ²	Tumour area (%)
C57BL/6 Rag2	2.0	44	3
	2.5	48	12
	5.0	92	25
	6.0	81	30
C57BL/6	2.0	5	<2
	2.0	<2	n.a.
	4.0	<2	n.a.
	4.0	<2	n.a.

Tumour number and area were determined in haematoxylin and eosin-stained histological sections of lungs (the total area examined for each mouse was 0.5–1.0 cm²) at the indicated times after vector exposure. Tumour areas were calculated from digitized images by using imaging software and are expressed as a percentage of the total lung area examined. Ignoring differences in the time of killing, tumour numbers in C57BL/6 Rag2 mice were significantly different ($P = 0.01$; two-tailed t -test) from those in C57BL/6 mice. n.a., not applicable.

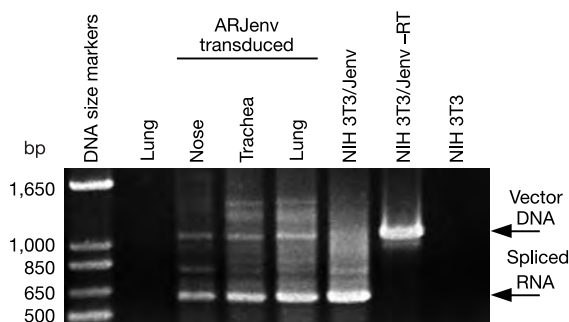


Figure 3 Env RNA expression in mouse lung and airways 4 months after the administration of ARJenv vector. RNA samples were subjected to reverse transcription and PCR amplification with primers flanking the intron in ARJenv (Fig. 1). The predicted product for spliced RNA is 597 base pairs (bp), whereas that for vector DNA is 1,065 bp. The 597-bp product was detected for all airway tissues of mice exposed to ARJenv vector and for Env-transformed NIH 3T3 mouse cells but was not detected for normal mouse lung. Amplification of residual vector DNA in RNA prepared from the cell line that was not treated with DNase or reverse transcriptase (–RT lane) showed the expected 1,065-bp vector DNA band.

months from two mice receiving the ARJenv AAV6 vector reduced the titre of the JSRV retroviral vector by 50% at a 1:3,000 dilution and completely neutralized the JSRV vector at a 1:100 dilution, whereas a 1:20 dilution of serum harvested at 2.5 months from a mouse that received only the ARAP4 vector (5×10^{10} vector genomes) had no effect on the JSRV vector titre. These results show that normal C57BL/6 mice can mount a strong serum antibody response against JSRV Env expressed from the AAV6 vector. Taken together, our results indicate that Env tumorigenesis can be controlled by the normal immune system of these mice.

The availability of antiserum against JSRV Env allowed us to examine JSRV Env expression in tumours from the immuno-deficient mice. Serum antibodies bound to tumours (Fig. 2k) and showed no binding to surrounding lung tissue (Fig. 2k) or to lungs of normal mice (not shown), indicating that the serum was specific for Env. Env staining was primarily localized to the cell membrane at the apical (airway) side of tumour cells (Fig. 2l), an appropriate position for retrovirus budding into the airway. This staining pattern was similar to that for SP-C, which is secreted into the airway, although SP-C was more diffusely localized to the cytoplasm near the apical membrane (Fig. 2i and data not shown). JSRV-infected sheep show increased secretion of virus-containing nasal fluid, a process facilitated by oncogenic transformation and expansion of the secretory alveolar type II cell population, and this is likely to be the main route for virus transmission. The level of Env staining was also correlated with tumour size (not shown), suggesting that higher levels of Env expression result in faster tumour growth. We were unable to detect Env staining of individual cells in the lung outside the tumours, whereas AP staining was clearly visible in such cells. This may indicate that Env expression was limited to alveolar type II cells or simply that our detection method was not sensitive enough to detect low-level Env expression in cells outside tumours.

Here we have shown that the Env protein of JSRV is sufficient to induce lethal adenocarcinoma in mice, with tumour appearance, location and alveolar type II pneumocyte marker staining similar to those found in sheep exposed to JSRV¹⁵. We found no significant homology of the JSRV Env protein to cellular proteins by database searches, arguing against an oncogenic mechanism involving the viral acquisition of part or all of a cellular oncogene. Because of the role of Hyal2 in human epithelial cell transformation¹¹ we were concerned that mice might not provide a suitable model for Env oncogenesis because mouse Hyal2 does not bind JSRV Env¹⁷. However, our more recent experiments showing that Madin–

Darby canine kidney (MDCK) epithelial cell transformation is dependent on the PI3K/Akt pathway and not the Hyal2 pathway⁸, and our inability to detect Hyal2 regulation of RON/Stk activity in fibroblasts¹⁸, encouraged us to test for Env tumorigenesis in mice. The fact that JSRV Env can induce tumours in mice shows that Env interaction with Hyal2 is not required for oncogenesis, although it may facilitate oncogenesis in other species.

AAV6 vectors promote efficient gene transfer and long-term gene expression in the lungs of young and adult animals. Their use provides some advantages over the use of transgenic mice engineered to study oncogenes, such as the ability to test new genes or combinations of genes rapidly, and the ability to introduce genes into a controllable number of cells at specific times without the possibility of leaky gene expression at earlier times during development. Avian retrovirus vectors have also been used in studies of lung cancer, but these require the use of transgenic mice that express the avian retrovirus receptor and, because the vectors infect only dividing cells, require infection of the lung *in utero*¹⁹.

The finding that JSRV Env can mediate the infection of human cells¹⁶ raised the possibility that JSRV or a related virus might cause cancer in humans. Furthermore, antibodies against the JSRV capsid proteins were found to react with about 30% of human bronchioalveolar carcinoma and pulmonary adenocarcinoma samples but not with adenocarcinomas from other organs or with normal tissues²⁰. Attempts to identify such viruses by polymerase chain reaction (PCR) have been unsuccessful so far^{21,22}. Results presented here show that immune responses against JSRV Env can limit oncogenesis in species other than sheep, and such immune responses may provide one level of protection against JSRV oncogenesis in humans. □

Methods

AAV6 vectors

AAV6 vectors expressing either the native (ARJenv) or a carboxy-terminal Flag-tagged (ARJenvF) JSRV Env were constructed by inserting the respective cDNAs and upstream splicing signals from pSX2-Jenv¹⁶ and pSX2neo-JenvFlag^{10,17} into the ARAP4 AAV vector¹³ in place of the AP cDNA (Fig. 1). ARJenv and ARJenvF are identical with the exception of the Flag sequence, which is present only in ARJenvF. ARJenvF had very poor transforming activity in cultured cells and was used as a negative control for tumorigenesis by the native Env protein. Virus was generated from the vector plasmids as described²³.

Vector administration to mice

Animal experiments were performed under guidelines approved by the Institutional Review Office of the Fred Hutchinson Cancer Research Center. Lightly anaesthetized mice received AAV vectors by the administration of drops to the nose, which were spontaneously inhaled. Mice killed at various times after vector administration were perfused by way of the heart with phosphate-buffered saline to remove blood from the lungs. Lungs were fixed in 2% paraformaldehyde in phosphate-buffered saline for 2.5 h at 22 °C and were washed with phosphate-buffered saline. Endogenous AP was inactivated by incubation at 68 °C for 1 h and lungs were stained for vector-encoded heat-stable AP as described²⁴.

Safety precautions

Vector production, vector administration to mice and harvesting of mouse tissues were performed in laminar-flow biological safety cabinets under BL2 containment using BL3 practices with careful attention to prevent aerosol generation. Workers were further protected by the use of particulate respirators. Agents determined to inactivate AAV vectors effectively (for example 1% bleach) were used for disinfection. Standard detergents and 70% ethanol were found to be relatively ineffective in the inactivation of AAV vectors.

Immunohistochemistry

Immunohistochemistry was performed on 4-μm sections of 2% paraformaldehyde-fixed paraffin-embedded tissues by following a standard avidin–biotin–peroxidase method. Cells with alveolar type II cell markers were detected with the use of two SP-C antibodies, Pro-C (provided by J. Whitsett) or anti-SP-C (Santa Cruz Biotechnology), and Clara cells were detected by using anti-CC10 antibodies (provided by J. Whitsett). After incubation with the primary antibodies and washing, the sections were sequentially incubated with biotinylated anti-rabbit IgG and the ABC-Elite reagent (Vector Labs). For immunohistochemical staining of tissues with anti-JSRV Env serum from ARJenv-transduced C57BL/6 mice, samples were incubated with unconjugated anti-mouse IgG and were then incubated with anti-JSRV Env serum or with control serum from an ARAP4-transduced mouse. After being washed, the sections were incubated sequentially

with biotinylated horse anti-mouse IgG and the ABC-Elite reagent. In all cases 3,3'-diaminobenzidine (with nickel chloride enhancement for SP-C and CC10 staining) was used as the chromagen and sections were counterstained with methyl green.

Analysis of Env RNA expression in lung and airways

Total RNA was isolated from lung, from trachea and from epithelial tissue scraped from the inside of the nose, by using a Polytron tissue homogenizer (Brinkmann) and Trizol RNA isolation reagent (Invitrogen). Samples were treated with DNase and with reverse transcriptase in the presence of a 3' Env primer; they were then subjected to 30 cycles of PCR amplification with primers flanking the intron in ARJenv (Fig. 1). Products were subjected to electrophoresis in agarose gels and were stained with ethidium bromide.

Received 22 December 2004; accepted 22 February 2005; doi:10.1038/nature03492.

1. Fan, H. (ed.) *Jaagsiekte Sheep Retrovirus and Lung Cancer* (Springer, Berlin, 2003).
2. Rai, S. K. et al. Candidate tumor suppressor HYAL2 is a glycosylphosphatidylinositol (GPI)-anchored cell-surface receptor for jaagsiekte sheep retrovirus, the envelope protein of which mediates oncogenic transformation. *Proc. Natl Acad. Sci. USA* **98**, 4443–4448 (2001).
3. Spencer, T. E., Mura, M., Gray, C. A., Griebel, P. J. & Palmerini, M. Receptor usage and fetal expression of ovine endogenous betaretroviruses: implications for coevolution of endogenous and exogenous retroviruses. *J. Virol.* **77**, 749–753 (2003).
4. Summers, C. et al. Systemic immune responses following infection with Jaagsiekte sheep retrovirus and in the terminal stages of ovine pulmonary adenocarcinoma. *J. Gen. Virol.* **83**, 1753–1757 (2002).
5. Sharp, J. M., Angus, K. W., Gray, E. W. & Scott, F. M. Rapid transmission of sheep pulmonary adenomatosis (jaagsiekte) in young lambs. Brief report. *Arch. Virol.* **78**, 89–95 (1983).
6. Maeda, N., Palmerini, M., Murgia, C. & Fan, H. Direct transformation of rodent fibroblasts by jaagsiekte sheep retrovirus DNA. *Proc. Natl Acad. Sci. USA* **98**, 4449–4454 (2001).
7. Allen, T. E. et al. The jaagsiekte sheep retrovirus envelope gene induces transformation of the avian fibroblast cell line DF-1 but does not require a conserved SH2 binding domain. *J. Gen. Virol.* **83**, 2733–2742 (2002).
8. Liu, S.-L. & Miller, A. D. Transformation of Madin–Darby canine kidney (MDCK) epithelial cells by sheep retrovirus envelope proteins. *J. Virol.* **79**, 927–933 (2004).
9. Palmerini, M. et al. A phosphatidylinositol 3-kinase docking site in the cytoplasmic tail of the Jaagsiekte sheep retrovirus transmembrane protein is essential for envelope-induced transformation of NIH 3T3 cells. *J. Virol.* **75**, 11002–11009 (2001).
10. Liu, S.-L., Lerman, M. I. & Miller, A. D. Putative phosphatidylinositol 3-kinase (PI3K) binding motifs in ovine betaretrovirus Env proteins are not essential for rodent fibroblast transformation and PI3K/Akt activation. *J. Virol.* **77**, 7924–7935 (2003).
11. Danilkovitch-Miagkova, A. et al. Hyaluronidase 2 negatively regulates RON receptor tyrosine kinase and mediates transformation of epithelial cells by jaagsiekte sheep retrovirus. *Proc. Natl Acad. Sci. USA* **100**, 4580–4585 (2003).
12. Halbert, C. L., Allen, J. M. & Miller, A. D. Adeno-associated virus type 6 (AAV6) vectors mediate efficient transduction of airway epithelial cells in mouse lungs compared to that of AAV2 vectors. *J. Virol.* **75**, 6615–6624 (2001).
13. Allen, J. M., Halbert, C. L. & Miller, A. D. Improved adeno-associated virus vector production with transfection of a single helper adenovirus gene, E4orf6. *Mol. Ther.* **1**, 88–95 (2000).
14. Malkinson, A. M. The genetic basis of susceptibility to lung tumors in mice. *Toxicology* **54**, 241–271 (1989).
15. Platt, J. A., Kraipowich, N., Villafane, F. & DeMartini, J. C. Alveolar type II cells expressing jaagsiekte sheep retrovirus capsid protein and surfactant proteins are the predominant neoplastic cell type in ovine pulmonary adenocarcinoma. *Vet. Pathol.* **39**, 341–352 (2002).
16. Rai, S. K., DeMartini, J. C. & Miller, A. D. Retrovirus vectors bearing jaagsiekte sheep retrovirus Env transduce human cells by using a new receptor localized to chromosome 3p21.3. *J. Virol.* **74**, 4698–4704 (2000).
17. Liu, S.-L., Duh, F.-M., Lerman, M. I. & Miller, A. D. Role of virus receptor Hyal2 in oncogenic transformation of rodent fibroblasts by sheep betaretrovirus env proteins. *J. Virol.* **77**, 2850–2858 (2003).
18. Miller, A. D., Van Hoven, N. S. & Liu, S.-L. Transformation and scattering activities of the receptor tyrosine kinase RON/Stk in rodent fibroblasts and lack of regulation by the jaagsiekte sheep retrovirus receptor, Hyal2. *BMC Cancer* **4**, 64 (2004).
19. Fisher, G. H. et al. Development of a flexible and specific gene delivery system for production of murine tumor models. *Oncogene* **18**, 5253–5260 (1999).
20. De las Heras, M. et al. Evidence for a protein related immunologically to the jaagsiekte sheep retrovirus in some human lung tumours. *Eur. Respir. J.* **16**, 330–332 (2000).
21. Yousem, S. A., Finkelstein, S. D., Swalsky, P. A., Bakker, A. & Otori, N. P. Absence of jaagsiekte sheep retrovirus DNA and RNA in bronchioloalveolar and conventional human pulmonary adenocarcinoma by PCR and RT-PCR analysis. *Hum. Pathol.* **32**, 1039–1042 (2001).
22. Hiatt, K. M. & Highsmith, W. E. Lack of DNA evidence for jaagsiekte sheep retrovirus in human bronchioloalveolar carcinoma. *Hum. Pathol.* **33**, 680 (2002).
23. Halbert, C. L., Allen, J. M. & Miller, A. D. Efficient mouse airway transduction following recombination between AAV vectors carrying parts of a larger gene. *Nature Biotechnol.* **20**, 697–701 (2002).
24. Halbert, C. L. et al. Transduction by adeno-associated virus vectors in the rabbit airway: efficiency, persistence, and readministration. *J. Virol.* **71**, 5932–5941 (1997).

Acknowledgements We thank J. C. DeMartini for providing histological pictures of human and sheep tumours shown in Fig. 2 and for discussions, and K. Hudkins-Loya and C. Alpers for help with the histological and antibody staining. This work was supported by NIH grants and a postdoctoral fellowship to S.K.W. from the National Sciences and Engineering Research Council of Canada.

Competing interests statement The authors declare that they have no competing financial interests.

Correspondence and requests for materials should be addressed to A.D.M. (dmiller@fhcc.org).

Activation of the DNA damage checkpoint and genomic instability in human precancerous lesions

Vassilis G. Gorgoulis^{1*}, Leandros-Vassilios F. Vassiliou^{1*}, Panagiotis Karakaidos¹, Panayotis Zacharatos¹, Athanassios Kotsinas¹, Triantafillos Liloglou², Monica Venere^{3,4}, Richard A. DiTullio Jr.^{3,4}, Nikolaos G. Kastrinakis¹, Brynny Levy⁶, Dimitris Kletsas⁷, Akihiro Yoneta³, Meenhard Herlyn³, Christos Kittas¹ & Thanos D. Halazonetis^{3,5}

¹Department of Histology and Embryology, School of Medicine, University of Athens, Athens GR-11527, Greece

²Roy Castle Lung Cancer Research Programme, Cancer Research Center, University of Liverpool, Liverpool L3 9TA, UK

³The Wistar Institute, Philadelphia, Pennsylvania 19104-4268, USA

⁴Graduate Group in Biomedical Sciences and

⁵Department of Pathology and Laboratory Medicine, University of Pennsylvania, Philadelphia, Pennsylvania 19104, USA

⁶Department of Human Genetics, Mount Sinai School of Medicine, New York, New York 10029, USA

⁷Institute of Biology, National Centre of Scientific Research 'Demokritos', Athens GR-15310, Greece

* These authors contributed equally to this work

DNA damage checkpoint genes, such as *p53*, are frequently mutated in human cancer, but the selective pressure for their inactivation remains elusive^{1–3}. We analysed a panel of human lung hyperplasias, all of which retained wild-type *p53* genes and had no signs of gross chromosomal instability, and found signs of a DNA damage response, including histone H2AX and Chk2 phosphorylation, *p53* accumulation, focal staining of *p53* binding protein 1 (53BP1) and apoptosis. Progression to carcinoma was associated with *p53* or 53BP1 inactivation and decreased apoptosis. A DNA damage response was also observed in dysplastic nevi and in human skin xenografts, in which hyperplasia was induced by overexpression of growth factors. Both lung and experimentally-induced skin hyperplasias showed allelic imbalance at loci that are prone to DNA double-strand break formation when DNA replication is compromised (common fragile sites). We propose that, from its earliest stages, cancer development is associated with DNA replication stress, which leads to DNA double-strand breaks, genomic instability and selective pressure for *p53* mutations.

The most frequently mutated gene in human cancer is *p53*, a gene that functions in the checkpoint response to DNA double-strand breaks (DSBs; Fig. 1a)^{1,2}. Several models (not mutually exclusive) have been proposed to explain the high frequency of *p53* inactivation^{3–6}. One of the prevailing models states that tumour growth leads to telomere attrition and hypoxia, resulting in a DNA damage response^{4–6}. This model predicts that the DNA damage response occurs some time after cancer initiation³. Here, we performed a systematic analysis of precancerous and cancer lesions to determine how early during human cancer development a DNA DSB checkpoint response might become apparent.

We first examined a previously described panel of surgically-resected, non-small cell lung carcinomas (NSCLCs) from patients who had received no form of cancer therapy before surgery^{7,8}. Almost all specimens in this panel ($n = 74$) contained normal adjacent lung tissue ($n = 72$), and some also contained hyperplastic ($n = 17$) and dysplastic lesions ($n = 2$), the location of which suggested that they were precursors of the NSCLCs. The *p53* gene was wild-type in all of the hyperplasias, mutant in the dysplasias, and either mutant ($n = 45$) or wild-type ($n = 29$) in the NSCLCs. For the two dysplasias, the same *p53* mutations were found in the adjacent NSCLCs, consistent with the dysplasias being precursors of the adjacent NSCLCs (data not shown).

The occurrence of a DNA damage response can be ascertained by monitoring histone H2AX phosphorylation (γ -H2AX), 53BP1 intracellular localization, Chk2 phosphorylation (on Thr68) and p53 protein levels^{9–13} (Fig. 1a). In the normal lung epithelium, all markers were consistent with the absence of a DNA damage response (Fig. 1b). However, in the hyperplasias there were signs that the DNA DSB checkpoint pathway had been activated. 53BP1, which is a sensor of DNA DSBs¹⁴, was localized to discrete nuclear foci. This is reminiscent of the 53BP1 foci that are observed in irradiated tissue culture cells and represent sites of DNA DSBs¹². In addition, histone H2AX and Chk2 were phosphorylated, p53 protein levels were elevated, and some cells were undergoing apoptosis (Fig. 1b). In the dysplasias and NSCLCs there were also signs of a DNA damage response, shown by 53BP1 localization and histone H2AX and Chk2 phosphorylation. However, apoptosis was

suppressed relative to the hyperplasias. In most cases this suppression was associated with p53 mutations or with low levels of wild-type p53 protein (Fig. 1b and 2). In a small number of NSCLCs, decreased apoptosis was associated with low levels of 53BP1 mRNA and protein (Fig. 1c, d and 2). In these NSCLCs, Chk2 was not phosphorylated on Thr68, but histone H2AX phosphorylation persisted (Fig. 1c and 2). These effects of low 53BP1 expression in NSCLCs replicate the effects of 53BP1 depletion in tissue culture cells^{13,15} (Supplementary Fig. 1).

To expand these studies to other tumour types, we examined a cohort of patients with malignant melanoma ($n = 61$). Eleven of these patients also had dysplastic nevi (nevi with melanocytic hyperplasia and atypia) adjacent to their melanoma. All of the dysplastic nevi, most of which were from areas of the body not exposed to sunlight, stained positive for phosphorylated histone

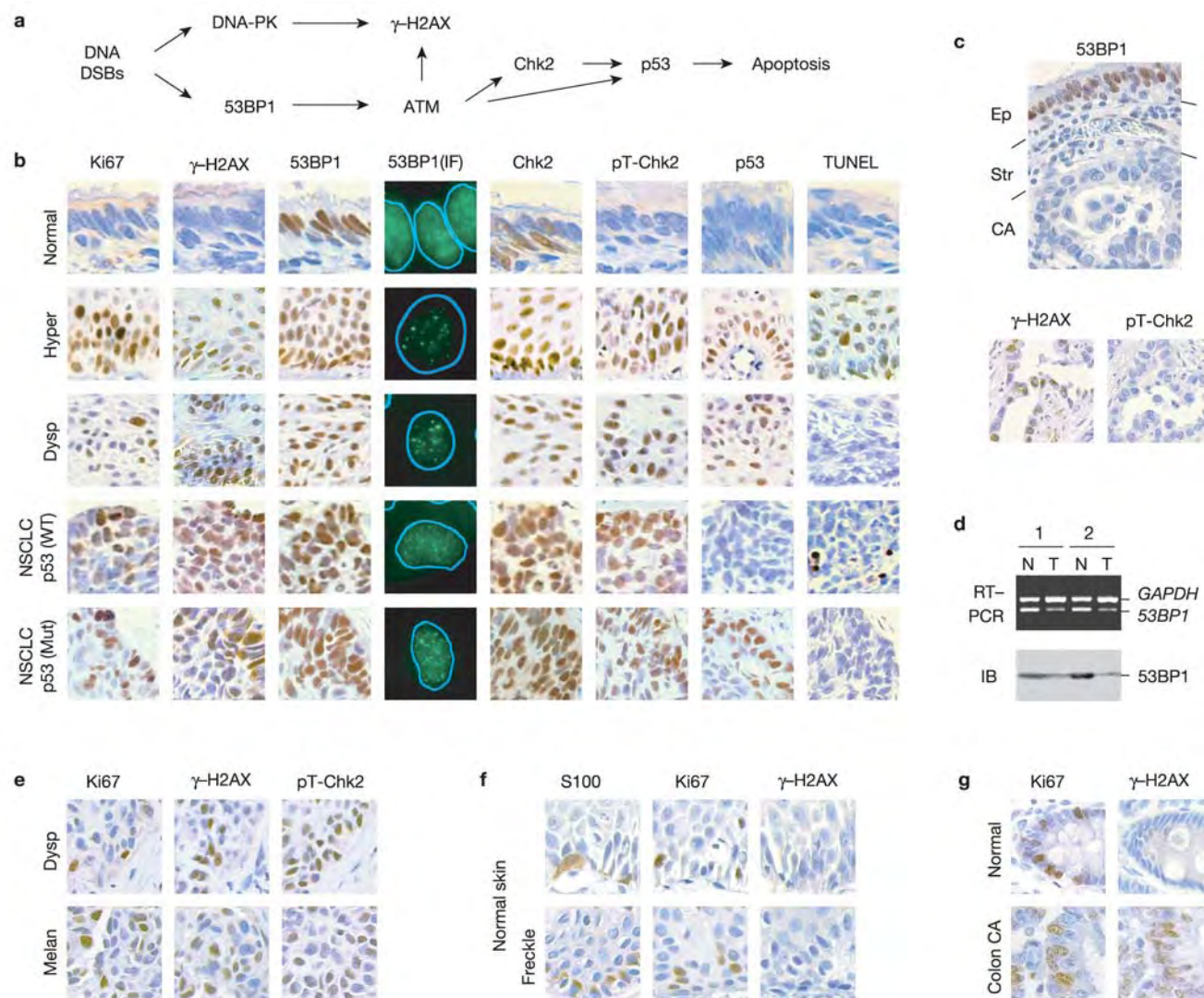


Figure 1 Activation of the DNA double-strand break (DSB) checkpoint pathway in human preneoplastic and neoplastic lesions. **a**, Current model of DNA DSB signalling pathway leading to p53-dependent apoptosis. For clarity, only a subset of the DNA DSB-response proteins are shown. DNA-PK, DNA-dependent protein kinase. **b**, Immunohistochemistry of normal bronchial epithelium, hyperplastic (Hyper), dysplastic (Dysp) and non-small cell lung carcinoma (NSCLC) lesions. p53(WT) and p53(Mut) are wild-type and mutant p53 genotypes, respectively. Ki67 and TUNEL staining were used to calculate proliferation and apoptotic indices, respectively. γ -H2AX, phosphorylated H2AX; pT-Chk2, Chk2 phosphorylated on Thr68. The distribution of 53BP1 in cells was examined by immunofluorescence (IF): 53BP1 is shown in green, nuclei are outlined in blue. **c**, Immunohistochemistry of a NSCLC that does not express 53BP1. Upper panel, a

section showing normal epithelium (Ep), stroma (Str) and NSCLC (CA) stained for 53BP1. Lower panels, adjacent NSCLC sections stained for γ -H2AX and pT-Chk2. **d**, Reverse transcriptase PCR (RT-PCR) and immunoblot (IB) analysis of two NSCLC lesions that stained negative for 53BP1 by immunohistochemistry. N, normal tissue; T, tumour tissue; GAPDH, glyceraldehyde-3-phosphate dehydrogenase. **e**, Immunohistochemistry of a dysplastic nevus (Dysp) and melanoma (Melan) from the same patient. **f**, Immunohistochemistry of two areas of normal skin adjacent to the dysplastic nevus shown in (e). One of the areas of normal skin (lower panels) corresponds to a freckle (simple lentigo). S100 staining marks the melanocytes. **g**, Immunohistochemistry of normal colonic epithelium and colon carcinoma (CA) from the same patient.

H2AX and Chk2 (Fig. 1e and 2), whereas the adjacent skin always stained negative (Fig. 1f). The melanomas also stained positive for phosphorylated histone H2AX and Chk2, with the exception of a small number of melanomas, which stained positive for γ -H2AX but negative for phosphorylated Chk2. These latter melanomas did not express either 53BP1 ($n = 5$) or Chk2 ($n = 6$) protein (Fig. 2). Unlike the lung hyperplasias, the percentage of apoptotic cells in the dysplastic nevi was low and did not decrease during progression to melanoma. Instead, progression to melanoma was associated with an increase in the proliferation index (Fig. 2). Although several models can be used to explain the increased proliferation that accompanies progression from dysplastic nevi to melanoma, one possibility is that activation of the DNA damage checkpoint in dysplastic nevi induces cell cycle arrest, and that progression to melanoma is associated with escape from this arrest through various mechanisms (including suppression of 53BP1 and Chk2 expression). In support of this model, dysplastic nevi stained positive for Cdc2 phosphorylated on Tyr 15 (Supplementary Fig. 2), which is a marker of G2 checkpoint activation^{2,11}.

The results presented above, as well as similar and complementary findings from another research group¹⁶, suggest that the DNA damage checkpoint is activated in a wide variety of human preneoplastic and neoplastic lesions. However, it is formally possible that the markers studied above would stain positive during normal cell proliferation, in which case the response of the preneoplastic and neoplastic lesions could simply be a reflection of their high proliferation index (Fig. 2). To explore this possibility we examined normal colonic crypts in surgically resected tissues from colon cancer patients. Despite having a higher proliferation index than the lung and melanocytic preneoplastic and neoplastic lesions, normal colonic epithelium ($n = 20$) stained uniformly negative for both histone H2AX and Chk2 phosphorylation (Fig. 1g and 2). Normal skin, which has a relatively high proliferation index, also stained negative (Fig. 1f and data not shown). Thus, activation of the DNA damage checkpoint occurs specifically in preneoplastic and neoplastic lesions.

The lung hyperplasias and dysplastic nevi that we studied are very early lesions in terms of the stage of cancer development, but they were probably precursors of the adjacent NSCLCs (see below) and melanomas, respectively. Because cancer can take years to develop, it is possible that the preneoplastic lesions did not show a DNA damage response when first formed. To address this issue we studied a model of hyperplasia¹⁷ in which human skin xenografts were implanted on the backs of immunodeficient mice and then induced to become hyperplastic by weekly subcutaneous injections (over 4 weeks) of adenoviral vectors expressing growth factors (basic fibroblast growth factor, stem cell factor and endothelin-3). Control xenografts were either not injected or injected with an adenovirus expressing green fluorescent protein (GFP). Both newborn human

foreskin and adult skin (from breast reduction cosmetic surgeries) were grafted. Foreskin from a specific donor, owing to its small size, was grafted on a single mouse, such that the various treatment groups consisted of grafts from different individuals. The larger adult skin samples were each used to prepare two grafts; one was injected with adenoviruses expressing growth factors, and the other served as a donor-matched untreated control. All grafts were harvested for analysis one week after the final injection.

Both the foreskin- and adult skin-derived hyperplastic xenografts mimicked the lung hyperplasias in terms of DNA damage response: 53BP1 was localized to discrete nuclear foci, histone H2AX and Chk2 were phosphorylated, p53 protein levels were induced and there was evidence of apoptosis (Fig. 3). None of these effects was observed in the non-injected controls or the controls injected with the adenovirus expressing GFP (Fig. 3e).

One possible mechanism for activation of the DNA damage checkpoint in the precancerous human lesions and hyperplastic skin xenografts might involve telomere attrition. We therefore compared telomere lengths in matched control and hyperplastic adult skin xenograft pairs. No differences between the groups were observed, but shortened telomeres were apparent in the K562 erythroleukaemia cancer cell line (Fig. 3d). The absence of telomere attrition, at least at this level of analysis, is perhaps not surprising given that the xenografts were obtained from newborns and young adults, and were examined just a few weeks after grafting.

Another possible mechanism for the DNA damage response observed in the precancerous lesions might involve replication stress. We noted high levels of cyclin E protein in both the lung hyperplasias and hyperplastic human skin xenografts (Supplementary Fig. 3 and data not shown). High levels of cyclin E, and more generally, deregulation of cyclin-dependent kinase activity in G1, are very frequent in human cancer, and compromise prereplication complex assembly and licensing of origins of replication. As a result, too few origins may fire or some origins may fire more than once per cell cycle. This would lead to replication stress, which in turn could lead to DNA DSBs, activation of the DNA damage checkpoint and even genomic instability^{18–22}.

As a first step in linking replication stress to the DNA damage response observed in preneoplasia and neoplasia, we studied cancer cell lines such as the Saos2 osteosarcoma and HeLa cervical carcinoma lines, in which the DNA damage checkpoint is active even in the absence of exposure to ionizing radiation¹⁵. If the DNA damage response in these cell lines is due to replication stress, it should be dependent on entry into S phase. Furthermore, the γ -H2AX foci in the nuclei of these cells should colocalize with ATR (ATM-Rad3-related protein)–ATRIP (ATR-interacting protein) foci because the ATR–ATRIP complex is recruited to sites of DNA replication stress^{23,24}. Both of these predictions were found to

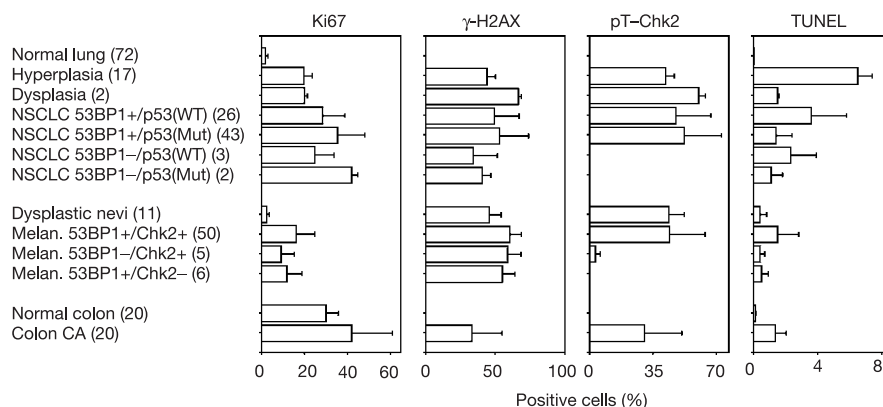


Figure 2 Summary of DNA double-strand break responses in human normal tissues, preneoplastic lesions and neoplastic lesions. The numbers in parentheses indicate the

number of samples examined. The bars show the mean $\pm$ s.d. of the percentage of cells that stained positive for the indicated marker.

be true. Saos2 cells stably transfected with a doxycycline-inducible *p21/WAF1/CDKN1A* gene were synchronized in G1 by inducing expression of *p21/WAF1/CDKN1A*, and then released into S phase by doxycycline withdrawal. Histone H2AX phosphorylation was more robust in cycling cells than in the cells arrested in G1, consistent with similar results by others²⁵. However, in early passage human diploid lung fibroblasts, histone H2AX was unphosphorylated, irrespective of whether the cells were cycling or resting (Supplementary Fig. 4a). Furthermore, a majority of γ -H2AX foci in non-irradiated Saos2 and HeLa cells were found to colocalize with ATRIP foci; as a control, γ -H2AX foci induced in these cell lines in response to ionizing radiation were devoid of ATRIP (Supplementary Fig. 4b).

To identify evidence for replication stress in preneoplastic human lesions, we reasoned that replication stress could lead to allelic imbalances, through formation of DNA DSBs and defective DNA repair. The loci in the genome that are prone to DNA DSB formation under conditions of replication stress are called common fragile sites^{26,27}. We predicted that these sites would be preferentially targeted for allelic imbalance in human preneoplastic lesions, and

tested this hypothesis using genomic DNA isolated from normal lung, hyperplastic and neoplastic tissue from 11 NSCLC patients. Allelic imbalance at the most common fragile site (*FRA3B* on chromosome 3p14) was compared to loci on chromosomes that commonly show allelic imbalance in advanced human cancers but do not correspond to common fragile sites. In the hyperplasias, allelic imbalance affecting the common fragile site *FRA3B* was very frequent (and occasionally extended to more distant loci on 3p), whereas the other chromosomal loci were either minimally affected or not affected (Fig. 4a). Consistent with preferential targeting of common fragile sites in these hyperplasias, comparative genomic hybridization analysis also revealed the absence of gross chromosomal instability (Supplementary Fig. 5 and data not shown). In the NSCLCs, allelic imbalances established in the hyperplastic stage persisted (providing evidence for progression from hyperplasia to NSCLC), but additional allelic imbalances affecting known tumour suppressor loci were also present, as was gross chromosomal instability (Fig. 4a, Supplementary Fig. 6 and data not shown). These and similar results^{16,28–30} are consistent with the presence of replication stress in human preneoplastic lesions.

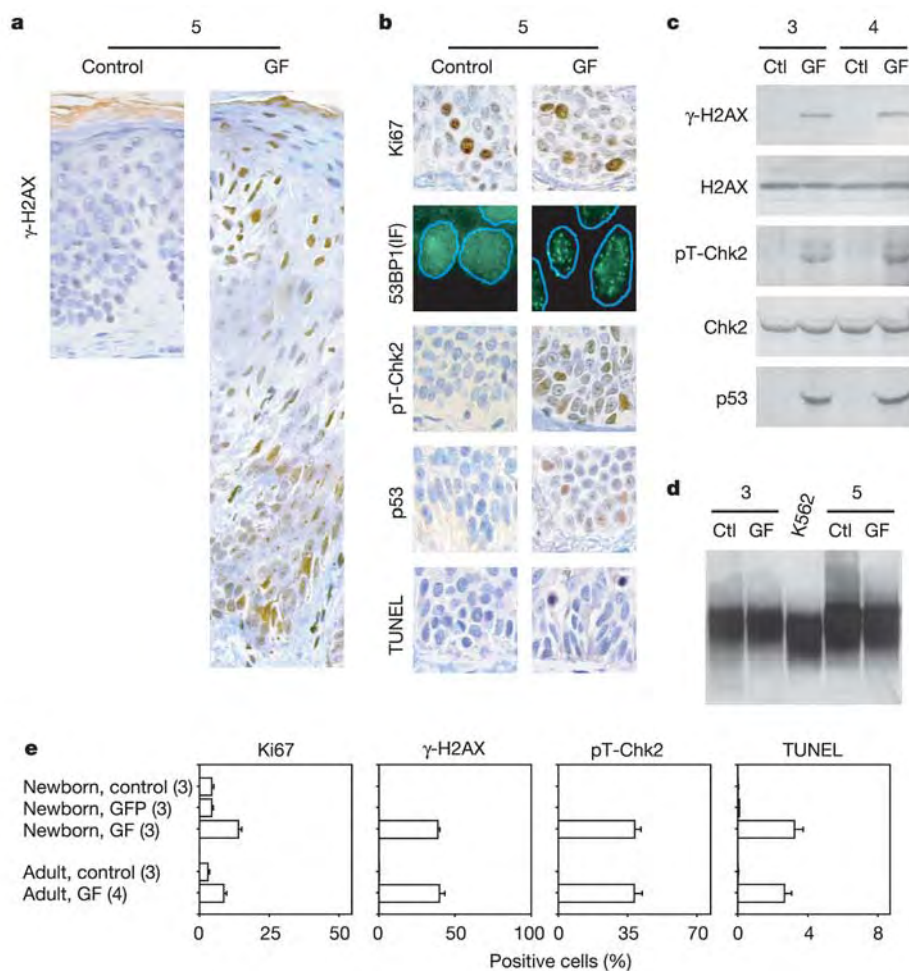


Figure 3 Activation of the DNA DSB checkpoint pathway in experimentally induced hyperplasias. Human newborn foreskin or adult skin xenografts were either untreated (control, Ctl), injected with an adenoviral vector that expresses GFP, or injected with adenoviral vectors overexpressing growth factors (GF). **a**, Cross-section of the entire epithelium of control and growth-factor-treated adult skin xenografts from the same donor (number 5), stained for γ -H2AX. **b**, Immunohistochemistry of control and growth factor-treated adult skin xenografts from donor 5 for Ki67, pT-Chk2 (Chk2 phosphorylated on Thr 68) and p53. TUNEL staining was used to mark the apoptotic cells and immunofluorescence (IF) was used to monitor the intracellular localization of 53BP1.

c, Immunoblot analysis of extracts prepared from control and growth factor-treated adult skin xenografts (from donors 3 and 4) to monitor H2AX and Chk2 phosphorylation (γ -H2AX and pT-Chk2, respectively) and p53 protein levels. **d**, Telomere length of control and growth-factor-treated adult skin xenografts from donors 3 and 5, assessed by Southern blot analysis. The erythroleukaemia cell line K562 was used as a control. **e**, Summary of DNA DSB responses in newborn foreskin and adult skin xenografts. Numbers in parentheses indicate the number of xenografts examined. The bars show the mean $\pm$ s.d. of the percentage of cells in each xenograft that stained positive for the indicated marker.

We extended the analysis of allelic imbalance at common fragile sites to the matched pairs of control and hyperplastic adult skin xenografts. Genomic DNA was isolated from 2–5 serial sections of xenograft tissue, and allelic imbalance was examined using two microsatellite markers (D3S1289 and D3S1300) that map to the *FRA3B* common fragile site on chromosome 3p, together with two markers (D3S1263 and D3S1566) that map to non-fragile sites on

chromosome 3p (3p25 and 3p13, respectively). In all three matched pairs examined, allelic imbalance at *FRA3B* was evident in the hyperplastic xenograft, whereas the 3p25 and 3p13 loci remained heterozygous (Fig. 4b). Thus, in an experimental human skin model, aberrant growth factor signalling associated with loss of tissue homeostasis (hyperplasia) can lead to allelic imbalance at common fragile sites within a few weeks.

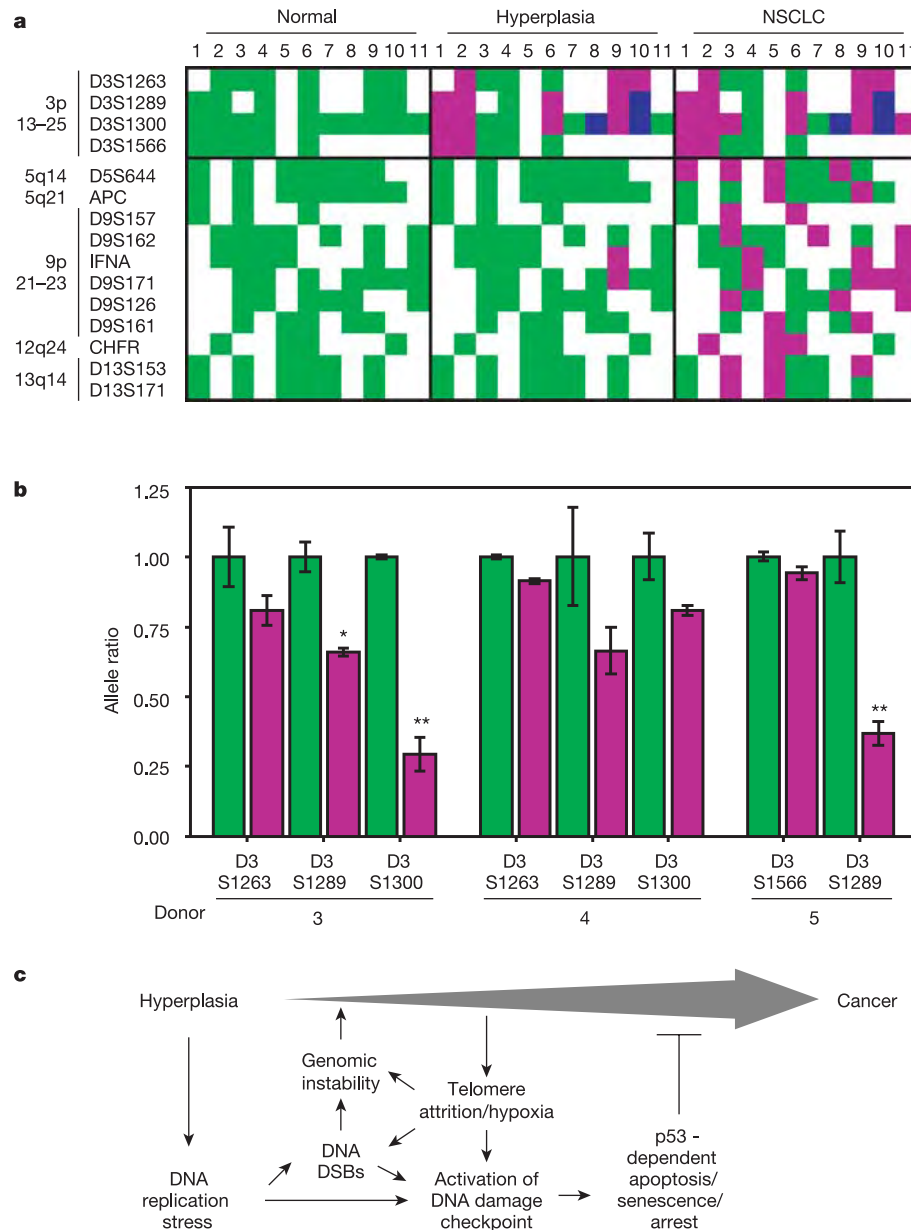


Figure 4 Allelic imbalance at common fragile sites in early human cancer lesions and model for activation of the DNA DSB checkpoint in cancer. **a**, Allelic imbalance at the common fragile site *FRA3B* on chromosome 3p14 is frequent in lung hyperplasias and precedes allelic imbalance at other loci. Allelic imbalance analysis of normal bronchial epithelium, hyperplastic and neoplastic tissue from 11 NSCLC patients. The microsatellite markers used in the analysis and their chromosomal position are indicated on the left. Markers D3S1289 and D3S1300 map to *FRA3B*; marker D3S1263 maps to 3p25; and marker D3S1566 maps to 3p13. Heterozygous loci (green), allelic imbalance (pink), microsatellite instability (blue), non-informative (homozygous) loci (white). **b**, Allelic imbalance at the common fragile site *FRA3B* in human adult skin xenografts. Data shown are means $\pm$ s.e.m. of allele ratios for microsatellite markers D3S1263, D3S1289, D3S1300 and D3S1566 from control (green) and growth-factor-treated (pink) xenografts from donors 3, 4 and 5. Ratios are not shown for non-informative (homozygous) loci. The

allele ratios were determined from two independent PCR reactions per sample and scaled so that the ratio would be 1 for the control xenografts. Asterisks indicate significantly different allele ratios between the control and growth factor-treated xenografts (single asterisk, $P < 0.005$; double asterisk, $P < 0.001$). In the matched pair from donor 4, the decrease in the allele ratio for marker D3S1289 in the hyperplastic xenograft reached statistical significance when tested using the mean variance, but not when tested using the individual variance. **c**, Model for activation of the DNA damage checkpoint in cancer. Aberrant stimulation of cell proliferation (as occurs in preneoplastic lesions) leads to DNA replication stress, DNA DSBs, genomic instability, activation of the DNA damage checkpoint, and p53-dependent apoptosis/senescence/arrest. p53-dependent apoptosis suppresses expansion of the precancerous lesion (*p53* tumour suppressor function) and provides selective pressure for *p53* inactivation. Telomere attrition and hypoxia can also contribute to DNA DSB formation, activation of the checkpoint and genomic instability.

Our findings, together with an accompanying study¹⁶, suggest a model that might explain the tumour suppressor function of *p53* and also hint to mechanisms leading to genomic instability in early cancer lesions (Fig. 4c). We propose that in precancerous lesions, aberrant stimulation of cell proliferation leads to DNA replication stress. This stress, either directly or through the formation of DNA DSBs, can activate the DNA damage checkpoint, which in turn induces cell cycle arrest or apoptosis, and thus functions as a tumour suppressor. In those cells that do not undergo permanent cell cycle arrest or apoptosis, there is a likelihood that errors in DNA DSB repair will lead to allelic imbalances. These imbalances will preferentially target common fragile sites because these sites are most sensitive to replication stress. At later stages, telomere attrition and hypoxia will also contribute to checkpoint activation and genomic instability^{4–6}. Eventually, tumour suppressor loci (such as *p53*) will be targeted, releasing the cells from the suppressive effects of the DNA damage checkpoint pathway and facilitating tumour progression. □

Methods

Antibodies

For immunohistochemistry, immunofluorescence and immunoblot analysis we used previously characterized primary antibodies^{13–15} at the indicated dilutions: anti-phospho-H2AX (Ser 139) (1:100; Upstate); anti-53BP1 and anti-Chk2 (hybridoma supernatants, 1:20; refs 13–15); anti-phospho-Chk2 (Thr 68, Lot 1) (1:100; Cell Signalling Technology); anti-p53 (DO7) (1:100; Dako); anti-Ki67 (MIB-1) (1:100; Dako); and anti-S100 (1:100; Dako).

Tissue samples

Our database of frozen and formalin-fixed, paraffin-embedded material from a total of 74 resected NSCLCs, and adjacent normal lung tissue and corresponding precancerous lesions (17 cases of hyperplasias/metaplasias, with two cases also bearing dysplasias) has previously been described^{7,8}. Sixty-one cases of sporadic malignant melanoma, 11 of which developed from dysplastic nevi, and 20 non-familial colon carcinoma cases were selected without bias from the patient population of the Agios Savas Hospital in Athens, Greece. None of the patients had undergone cancer therapy before surgical resection of the lesions.

Human skin xenograft model

The human skin xenograft model has been described¹⁷. For these studies, we examined 9 newborn foreskin xenografts: 3 were injected subcutaneously with adenoviruses expressing basic fibroblast growth factor, stem cell factor and endothelin-3; 3 were injected with an adenovirus expressing GFP; and 3 were left untreated. Injections were weekly over a period of 4 weeks. We also studied 4 adult skin grafts from patients undergoing breast reduction cosmetic surgeries. These grafts were cut in half, and each half was implanted in a separate SCID mouse. For each graft, one mouse received injections with adenoviral vectors expressing the growth factors listed above, and in the other mouse the graft was untreated. All grafts were harvested a week after the last injection. The foreskin grafts were fixed with formalin and analysed by immunohistochemistry. Part of each adult skin graft was flash-frozen and used to prepare protein extracts, and another part was fixed with formalin and analysed by immunohistochemistry or used to prepare genomic DNA. One of the adult untreated grafts was damaged during sectioning and was not analysed by immunohistochemistry.

Analysis of tissue samples

Formalin-fixed tissue sections were processed for immunohistochemistry, immunofluorescence and Tdt-mediated dUTP nick-end labelling (TUNEL) analysis as previously described^{7,8}. RNA was extracted from frozen samples and used to prepare complementary DNA^{7,8}. For analysis of 53BP1 (also known as TP53BP1) expression at the mRNA level, the cDNA was amplified by semiquantitative multiplex polymerase chain reaction (PCR) using primers specific for 53BP1 (forward primer 3'-GCAGCCTCTGTGAA GCAGCA-5'; reverse 3'-ATGCCAAGGAATCCAGTTACACACAA-5') and GAPDH (GAPD), as standard^{7,8}. *p53* mutations were identified by single-strand conformation polymorphism analysis and by sequencing^{7,8}. Proteins were extracted by lysis of minced frozen samples using RIPA buffer supplemented with protease inhibitors¹³. Histones were isolated from the RIPA-insoluble pellet by extraction with buffer consisting of 10 mM HEPES, 1.5 mM MgCl₂, 10 mM KCl, 0.5 mM dithiothreitol, 1.5 mM phenylmethyl sulphonyl fluoride and 0.25 N HCl for 1 h at 4 °C. Immunoblotting was performed as previously described¹³.

Telomere length assay

Telomere lengths were determined using the TeloTAGGG Telomere Length Assay (Roche Diagnostics) according to the manufacturer's instructions. Briefly, 4 µg genomic DNA isolated from two matched pairs of control and growth factor-treated adult skin xenografts was digested with *HinfI* and *RsaI*, and subjected to Southern blot analysis using telomere-specific labelled probes. DNA prepared from the K562 erythroleukaemic cell line served as a control.

Allelic imbalance analysis

Allelic imbalance analysis of lung tissues was tabulated from our previously published studies (refs 7, 8 and references therein), or extended to include microsatellite markers for chromosome 3p. For analysis of human skin xenografts, 2–5 serial 10 µm-thick paraffin-embedded sections were microdissected using laser capture, and genomic DNA was extracted as previously described⁷. Each genomic DNA sample was subjected to two independent PCR reactions, and the PCR products were resolved using a 377 ABI PRISM automated sequencer (Applied Biosystems), as previously described⁸. Differences in allele ratios between control and growth factor-treated xenografts were evaluated twice for statistical significance (using the individual variances and the mean variance calculated from all the replicates), and were scored positive only if both evaluations showed statistical significance.

Received 15 December 2004; accepted 18 February 2005; doi:10.1038/nature03485.

- Hollstein, M., Sidransky, D., Vogelstein, B. & Harris, C. C. *p53* mutations in human cancers. *Science* **253**, 49–53 (1991).
- Kastan, M. B. & Bartek, J. Cell-cycle checkpoints and cancer. *Nature* **432**, 316–323 (2004).
- Halazonetis, T. D. Constitutively active DNA damage checkpoint pathways as the driving force for the high frequency of *p53* mutations in human cancer. *DNA Repair (Amst.)* **3**, 1057–1062 (2004).
- Takai, H., Smogorzewska, A. & de Lange, T. DNA damage foci at dysfunctional telomeres. *Curr. Biol.* **13**, 1549–1556 (2003).
- d'Adda di Fagagna, F. *et al.* A DNA damage checkpoint response in telomere-initiated senescence. *Nature* **426**, 194–198 (2003).
- Graeber, T. G. *et al.* Hypoxia-mediated selection of cells with diminished apoptotic potential in solid tumours. *Nature* **379**, 88–91 (1996).
- Gorgoulis, V. G. *et al.* Alterations of the p16-pRb pathway and the chromosome locus 9p21–22 in non-small-cell lung carcinomas: relationship with *p53* and MDM2 protein expression. *Am. J. Pathol.* **153**, 1749–1765 (1998).
- Karakaidos, P. *et al.* Overexpression of the replication licensing regulators hCdt1 and hCdc6 characterizes a subset of non-small-cell lung carcinomas: synergistic effect with mutant *p53* on tumor growth and chromosomal instability—evidence of E2F-1 transcriptional control over hCdt1. *Am. J. Pathol.* **165**, 1351–1365 (2004).
- Kastan, M. B., Onyekwere, O., Sidransky, D., Vogelstein, B. & Craig, R. W. Participation of *p53* protein in the cellular response to DNA damage. *Cancer Res.* **51**, 6304–6311 (1991).
- Rogakou, E. P., Pilch, D. R., Orr, A. H., Ivanova, V. S. & Bonner, W. M. DNA double-stranded breaks induce histone H2AX phosphorylation on serine 139. *J. Biol. Chem.* **273**, 5858–5868 (1998).
- Bartek, J. & Lukas, J. Chk1 and Chk2 kinases in checkpoint control and cancer. *Cancer Cell* **3**, 421–429 (2003).
- Schultz, L. B., Chehab, N. H., Malikzay, A. & Halazonetis, T. D. *p53* binding protein 1 (53BP1) is an early participant in the cellular response to DNA double-strand breaks. *J. Cell Biol.* **151**, 1381–1390 (2000).
- Mochan, T. A., Venere, M., DiTullio, R. A. Jr & Halazonetis, T. D. 53BP1 and NFB1/MDC1-Nbs1 function in parallel interacting pathways activating ataxia-telangiectasia mutated (ATM) in response to DNA damage. *Cancer Res.* **63**, 8586–8591 (2003).
- Huyen, Y. *et al.* Methylated lysine 79 of histone H3 targets 53BP1 to DNA double-strand breaks. *Nature* **432**, 406–411 (2004).
- DiTullio, R. A. Jr *et al.* 53BP1 functions in an ATM-dependent checkpoint pathway that is constitutively activated in human cancer. *Nature Cell Biol.* **4**, 998–1002 (2002).
- Bartkova, J. *et al.* DNA damage response as a candidate anti-cancer barrier in early human tumorigenesis. *Nature* doi:10.1038/nature03482 (this issue).
- Berking, C. *et al.* Induction of melanoma phenotypes in human skin by growth factors and ultraviolet B. *Cancer Res.* **64**, 807–811 (2004).
- Spruck, C. H., Won, K. A. & Reed, S. I. Deregulated cyclin E induces chromosome instability. *Nature* **401**, 297–300 (1999).
- Lengronne, A. & Schwob, E. The yeast CDK inhibitor Sic1 prevents genomic instability by promoting replication origin licensing in late G1. *Mol. Cell* **9**, 1067–1078 (2002).
- Tanaka, S. & Diffley, J. F. Deregulated G1-cyclin expression induces genomic instability by preventing efficient pre-RC formation. *Genes Dev.* **16**, 2639–2649 (2002).
- Vaziri, C. *et al.* A *p53*-dependent checkpoint pathway prevents rereplication. *Mol. Cell* **11**, 997–1008 (2003).
- Eklholm-Reed, S. *et al.* Deregulation of cyclin E in human cells interferes with prereplication complex assembly. *J. Cell Biol.* **165**, 789–800 (2004).
- Tibbetts, R. S. *et al.* Functional interactions between BRCA1 and the checkpoint kinase ATR during genotoxic stress. *Genes Dev.* **14**, 2989–3002 (2000).
- Cortez, D., Guntuku, S., Qin, J. & Elledge, S. J. ATR and ATRIP: partners in checkpoint signaling. *Science* **294**, 1713–1716 (2001).
- MacPail, S. H., Banath, J. P., Yu, Y., Chu, E. & Olive, P. L. Cell cycle-dependent expression of phosphorylated histone H2AX: reduced expression in unirradiated but not X-irradiated G1-irradiated G1-phase cells. *Radiat. Res.* **159**, 759–767 (2003).
- Arlt, M. F., Casper, A. M. & Glover, T. W. Common fragile sites. *Cytogenet. Genome Res.* **100**, 92–100 (2003).
- Casper, A. M., Nghiem, P., Arlt, M. F. & Glover, T. W. ATR regulates fragile site stability. *Cell* **111**, 779–789 (2002).
- Mao, L. *et al.* Frequent microsatellite alterations at chromosomes 9p21 and 3p14 in oral premalignant lesions and their value in cancer risk assessment. *Nature Med.* **2**, 682–685 (1996).
- Wistuba, I. I. *et al.* High resolution chromosome 3p allelotyping of human lung cancer and preneoplastic/preinvasive bronchial epithelium reveals multiple, discontinuous sites of 3p allele loss and three regions of frequent breakpoints. *Cancer Res.* **60**, 1949–1960 (2000).
- Maitra, A. *et al.* High-resolution chromosome 3p allelotyping of breast carcinomas and precursor lesions demonstrates frequent loss of heterozygosity and a discontinuous pattern of allele loss. *Am. J. Pathol.* **159**, 119–130 (2001).

Supplementary Information accompanies the paper on www.nature.com/nature.

Acknowledgements The authors thank R. Kaufman, L. Kaklamanis, M. Arnaouti, P. Foukas, K. Ryan and V. Kostaki for support, reagents and tissue samples. This work was supported by grants to T.D.H. from the National Cancer Institute and to T.L. from the Roy Castle Lung Foundation, UK. M.V. was supported by a Radiation training grant from the NIH.

Competing interests statement The authors declare that they have no competing financial interests.

Correspondence and requests for materials should be addressed to T.D.H. (halazonetis@wistar.upenn.edu).

Specific killing of BRCA2-deficient tumours with inhibitors of poly(ADP-ribose) polymerase

Helen E. Bryant¹, Niklas Schultz², Huw D. Thomas³, Kayan M. Parker¹, Dan Flower¹, Elena Lopez¹, Suzanne Kyle³, Mark Meuth¹, Nicola J. Curtin³ & Thomas Helleday^{1,2}

¹The Institute for Cancer Studies, University of Sheffield, Medical School, Beech Hill Road, Sheffield S10 2RX, UK

²Department of Genetics, Microbiology and Toxicology, Arrhenius Laboratory, Stockholm University, S-106 91 Stockholm, Sweden

³Northern Institute for Cancer Research, University of Newcastle upon Tyne, Medical School, Newcastle upon Tyne, NE2 4HH, UK

Poly(ADP-ribose) polymerase (PARP1) facilitates DNA repair by binding to DNA breaks and attracting DNA repair proteins to the site of damage^{1–3}. Nevertheless, PARP1^{−/−} mice are viable, fertile and do not develop early onset tumours⁴. Here, we show that PARP inhibitors trigger γ -H2AX and RAD51 foci formation. We

propose that, in the absence of PARP1, spontaneous single-strand breaks collapse replication forks and trigger homologous recombination for repair. Furthermore, we show that BRCA2-deficient cells, as a result of their deficiency in homologous recombination, are acutely sensitive to PARP inhibitors, presumably because resultant collapsed replication forks are no longer repaired. Thus, PARP1 activity is essential in homologous recombination-deficient BRCA2 mutant cells. We exploit this requirement in order to kill BRCA2-deficient tumours by PARP inhibition alone. Treatment with PARP inhibitors is likely to be highly tumour specific, because only the tumours (which are BRCA2^{−/−}) in BRCA2^{+/−} patients are defective in homologous recombination. The use of an inhibitor of a DNA repair enzyme alone to selectively kill a tumour, in the absence of an exogenous DNA-damaging agent, represents a new concept in cancer treatment.

Despite its important role in the cellular response to genotoxic stress, PARP1 is not required for survival in the absence of such an insult, and PARP1^{−/−} mice are viable and fertile^{2,5,6}. These mice do not develop early onset tumours and tumour latency is increased in PARP1 knockout mice that are deficient for p53 (ref. 4). Nevertheless, it is generally accepted that loss of PARP1 activity is important in maintaining genetic stability, because PARP1^{−/−} mice exhibit defective DNA single-strand break (SSB) repair and an increase in homologous recombination, sister chromatid exchange and micronuclei formation^{1,2,5–7}. However, the elevated homologous recombination levels in PARP1^{−/−} mice represent an error-free repair pathway, which may explain why the genetic instability in PARP1-deficient or inhibited cells is not associated with any accumulation of mutations or cancer.

PARP1 does not seem to be directly involved in homologous recombination, as RAD51 foci form normally in PARP1-deficient cells and homologous recombination-mediated repair of a DNA double-strand break (DSB) is unaffected by inhibition or loss of

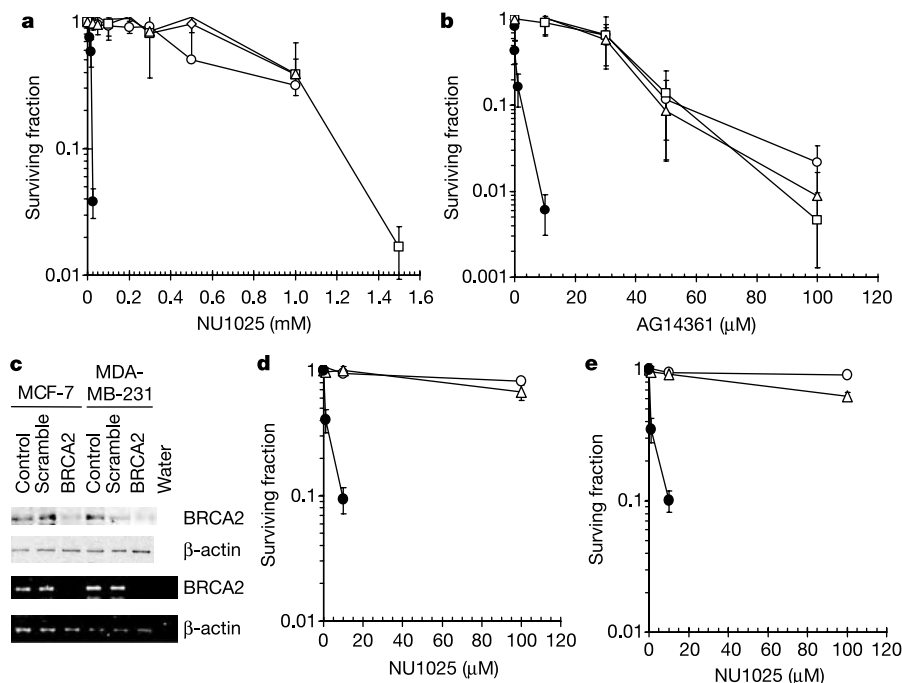


Figure 1 BRCA2-deficient cells are hypersensitive to inhibitors of PARP. **a**, **b**, Colony outgrowth of V79 (wild type; open circles), V-C8 (BRCA2-deficient¹⁴; filled circles), V-C8#13 (V-C8 complemented with *BRCA2* on human chromosome 13 (ref. 14); squares) and V-C8+B2 (V-C8 complemented with *BRCA2* on an expression vector¹⁴; triangles) cells upon continuous exposure to the PARP inhibitor NU1025 (**a**) or after a 24-h treatment with AG14361 (**b**). **c**, Western blot and RT-PCR analysis of protein and RNA lysates

isolated from MCF-7 (p53^{WT}) or MDA-MB-231 (p53^{mut}) breast cancer cells after 48-h transfection with siRNA. **d**, **e**, Clonogenic survival of siRNA-treated MCF-7 cells (**d**) or MDA-MB-231 cells (**e**) after exposure to the PARP inhibitor NU1025. Open circles, control; triangles, scramble siRNA; filled circles, BRCA2 siRNA. The means and standard deviation of at least three experiments are shown.

PARP1 (refs 8, 9). Therefore, increased homologous recombination levels are most likely explained by an accumulation of recombinogenic substrates in PARP1-deficient cells. Thus, homologous recombination may have a particularly important role in repairing lesions in these cells, and we propose that homologous recombination-deficient cells may be especially sensitive to inhibition of PARP1.

To test our hypothesis, we treated the *irs1* and *irs1SF* cell lines (defective in *XRCC2* (ref. 10) or *XRCC3* (ref. 11), respectively) with increasing doses of the PARP inhibitors 3-aminobenzamide (3-AB), 1,5-dihydroxyisoquinoline (ISQ), 8-hydroxy-2-methylquinazolinone (NU1025, ref. 12) or 1-(4-dimethylaminomethyl-phenyl)-8,9-dihydro-7H-2,7,9a-benzo[cd]azulen-6-one (AG14361, ref. 13). Both of the homologous recombination-deficient cell lines were sensitive to all PARP inhibitors, and correction of the homologous recombination defect with a functional *XRCC2* or *XRCC3* gene reversed the sensitivity (Supplementary Fig. S1). This supports our hypothesis that recombinogenic lesions spontaneously accumulate in the absence of PARP1 activity and that these are lethal in the absence of homologous recombination.

We next determined whether cells defective in homologous recombination as a result of non-functional *BRCA2* are also sensitive to the cytotoxicity of PARP inhibitors. We discovered that the potent specific PARP inhibitors NU1025 and AG14361 were profoundly cytotoxic at low concentrations in the *BRCA2*-deficient cell line V-C8 (ref. 14), as compared with the V79 wild-type cell line

or the V-C8 cell line complimented with *BRCA2* (ref. 14) (Fig. 1a, b). Thus, the sensitivity to the PARP inhibitor is a direct consequence of the *BRCA2* defect. This is also supported by sensitivity of both *BRCA1*^{-/-} and *BRCA2*^{-/-} embryonic stem (ES) cells to the PARP inhibitors KU0058684 and KU0058948 (ref. 15).

BRCA2 defects are associated with familial breast cancer¹⁶. To examine the effect of PARP inhibition in a more relevant model we examined whether the MCF7 (wild-type p53) and MDA-MB-231 (mutated p53) human breast cancer cell lines displayed a similar sensitivity to NU1025 upon depletion of *BRCA2* with a mixture of *BRCA2* short interfering RNAs (siRNA) (Fig. 1c). We found that PARP inhibitors profoundly reduced the survival of MCF7 and MDA-MB-231 cells only when *BRCA2* was depleted (Fig. 1d, e). This shows that *BRCA2*-depleted breast cancer cells are sensitive to PARP inhibitors regardless of p53 status.

The specific killing of *BRCA2*-deficient cells with NU1025 or AG14361 is not a secondary effect due to a differential sensitivity of the cells to the inhibitors, because PARP1 activity was equally inhibited by NU1025 or AG14361 regardless of the *BRCA2* status (Supplementary Fig. S2).

There are two DNA damage-activated PARPs present in the nucleus of mammalian cells—PARP1 and PARP2—and all reported PARP inhibitors inhibit both. In order to distinguish which PARP was responsible for the effect, we tested whether the absence of PARP1 and/or PARP2 results in accumulation of toxic lesions, by depleting these and *BRCA2* with siRNA in human cells (Fig. 2a). We found that clonogenic survival was significantly reduced when both PARP1 and *BRCA2* proteins were co-depleted from human cells (Fig. 2b). Depletion of PARP2 with *BRCA2* had no effect on the clonogenic survival, and depletion of PARP2 in PARP1- and *BRCA2*-depleted cells did not result in additional toxicity. These results suggest that PARP1 rather than PARP2 is responsible for protection against toxic recombinogenic lesions spontaneously occurring in human cells. The cloning efficiency was only reduced by 40% compared with control in PARP1 and *BRCA2* co-depleted cells, while most homologous recombination-deficient cells died after treatments with PARP inhibitors at concentrations that were non-toxic to normal cells. This may be attributable to incomplete depletion of the abundant PARP1 protein by siRNA (Fig. 2c), which might be sufficient to maintain PARP1 function in some of the cells. Alternatively, inhibition of PARP1 activity, without affecting its DNA-binding activity, may, by hindering the dissociation of PARP1 from DNA, result in a more serious lesion. Depletion of PARP1 and PARP2 resulted in lower clonogenic survival, in agreement with the embryonic lethal phenotype of PARP1 and PARP2 double knockout mice¹⁷.

About 10⁴ spontaneous SSBs occur in each cell every day¹⁸; in the absence of PARP1 activity these SSBs may be converted to DSBs during replication and therefore collapse replication forks. To determine whether DSBs are formed after treatment with NU1025, we investigated the formation of γ -H2AX foci, which normally form in response to DSBs in mammalian cells¹⁹(Fig. 3a, b).

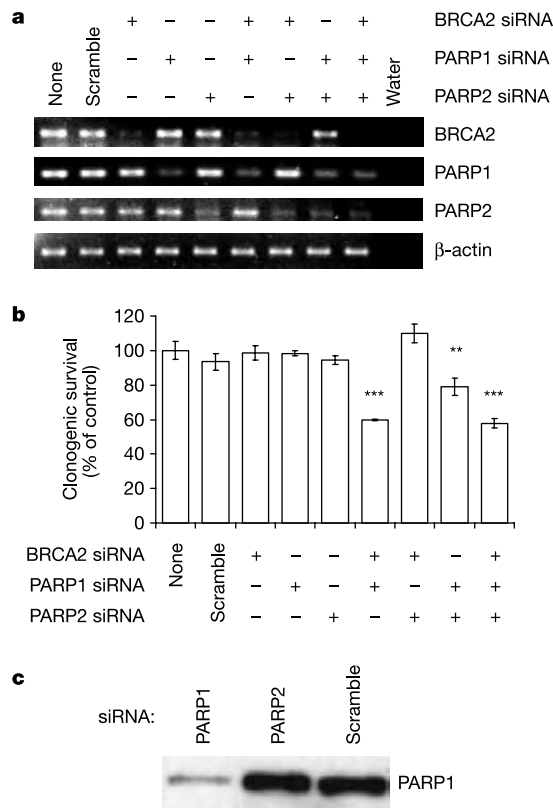


Figure 2 PARP1 and not PARP2 is important in preventing formation of a recombinogenic lesion, causing death in the absence of *BRCA2*. **a**, RT-PCR on RNA isolated from human SW480SN.3 cancer cells treated with *BRCA2*, *PARP1* and *PARP2* siRNA in combinations, as shown, for 48 h. **b**, Clonogenic survival after 48 h depletion of *BRCA2*, *PARP1* and *PARP2*. The means and standard deviation of at least three experiments are shown. Two and three asterisks designate statistical significance for *t*-test at *P* < 0.01 and *P* < 0.001, respectively. **c**, Western blot for PARP1 in SW480SN.3 cells subjected to different siRNA treatments.

Table 1 Response of V-C8 and V-C8 + B2 cells to treatment with AG14361 in tumour xenografts

Cell line	Treatment	Number responded/ total number of mice with tumours	Detail
V-C8+B2	Control	0/4	0
V-C8+B2	25 mg ml ⁻¹ AG14361	0/3	0
V-C8+B2	50 mg ml ⁻¹ AG14361	0/2	0
V-C8	Control	0/5	0
V-C8	25 mg ml ⁻¹ AG14361	2/3	1 CR, 1MR
V-C8	50 mg ml ⁻¹ AG14361	1/2	1 PR

CR, complete response, total regression of tumour, no tumour at autopsy; MR, minor response, 25% reduction in tumour; PR, partial response, 50% reduction in tumour.

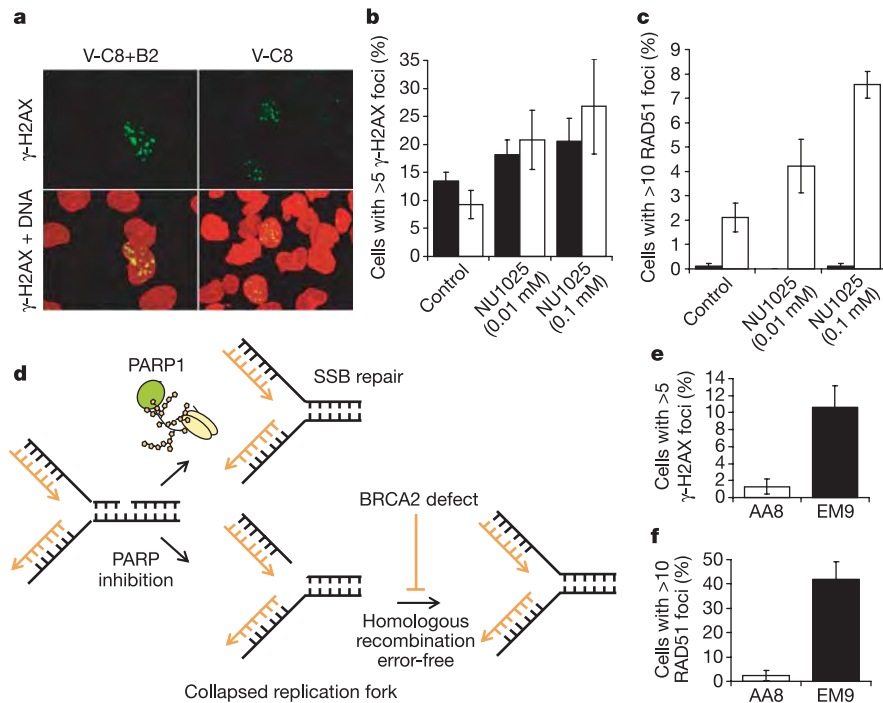


Figure 3 BRCA2-deficient cells fail to repair a recombination lesion formed by inhibitors of PARP. **a**, Visualization of γ -H2AX foci in untreated V-C8+B2 and V-C8 cells. **b**, **c**, Number of cells containing γ -H2AX foci (**b**) or RAD51 foci (**c**) visualized in V-C8+B2 (open columns) and V-C8 (filled columns) cells after a 24-h treatment with NU1025 (10 μ M). **d**, Model: in the absence of PARP1, unrepaired SSBs collapse replication forks into

DSBs, which are repaired by BRCA2-dependent homologous recombination.

e, **f**, Spontaneous γ -H2AX foci (**e**) or RAD51 foci (**f**) in wild-type AA8 cells or SSB repair-deficient EM9 cells lacking XRCC1. The means and standard deviation of three to nine experiments are shown.

We found an induction of γ -H2AX foci in both V-C8 and V-C8+B2 (BRCA2 complemented) cells treated with NU1025 (statistically significant in *t*-test, $P < 0.05$; Fig. 3b).

Replication-associated DSBs are normally repaired by the error-free homologous recombination pathway in mammalian cells²⁰. In the absence of BRCA2 and homologous recombination these collapsed replication forks may persist and be lethal (Fig. 3d). To test whether this is the case in the BRCA2-deficient V-C8 cells we determined the ability of these and BRCA2 complemented cells to form RAD51 foci in response to NU1025. We found that RAD51 foci were indeed induced in V-C8+B2 cells after treatment with NU1025 (statistically significant in *t*-test, $P < 0.05$; Fig. 3c). This indicates that recombinogenic lesions (probably collapsed replication forks) trigger homologous recombination repair in these cells, allowing them to survive. In contrast, the BRCA2-deficient V-C8 cells were unable to form RAD51 foci in response to NU1025 treatment (Fig. 3c), confirming the homologous recombination defect, which results in unrepaired recombinogenic lesions and cell death (Fig. 3d).

If this model is correct, we predict that cells deficient in other SSB repair proteins would have more spontaneously collapsed forks (more γ -H2AX foci), which would, in turn, trigger more homologous recombination (more RAD51 foci). To test this, we determined γ -H2AX and RAD51 foci in the SSB-repair-deficient EM9 cell line (with a defect in XRCC1 (ref. 21)). We found an increase in both spontaneous γ -H2AX and RAD51 foci in EM9 cells (statistically significant in *t*-test, $P < 0.05$; Fig. 3e, f), which supports our hypothesis that lack of SSB repair causes the recombinogenic lesions formed in PARP-inhibited cells. Furthermore, our model explains the increased level of sister chromatid exchange generally found in both PARP1- and XRCC1-deficient cells^{5,6,21}.

Familial breast cancer is commonly caused by an inherited defect in one of the BRCA1 or BRCA2 alleles¹⁶. During the course of life,

the functional BRCA1 or BRCA2 allele can be lost in some cells, thereby initiating the development of a tumour. Thus, the tumours that arise are BRCA1 or BRCA2 deficient (BRCA2^{-/-}) whereas the remaining somatic cells still have functional BRCA proteins (BRCA2^{+/+}). If we exploit the fact that the PARP1 pathway becomes essential only in homologous recombination-deficient cells, we could develop a treatment highly specific for the BRCA2-deficient tumour cells. Furthermore, PARP inhibitors are likely to produce few side effects, because PARP1 knockout mice are viable and healthy^{5,6}, and prolonged PARP1 inhibition is tolerated in mice bearing tumour xenografts after the administration of AG14361

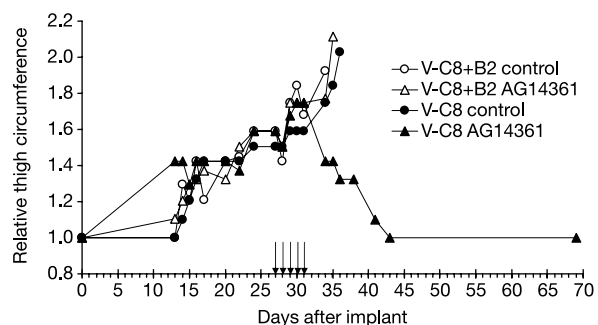


Figure 4 Specific killing of a BRCA2 tumour after a 5-day treatment with a PARP1 inhibitor in xenografts. Growth of V-C8 or V-C8+B2 tumours injected intramuscularly into the thigh of mice after treatment with AG14361 (25 mg kg⁻¹ in 10 ml kg⁻¹ saline) or with saline (10 ml kg⁻¹) alone. Tumour growth is measured as relative difference in diameter between the injected and non-injected leg. The tumour growth from one mouse in each treatment group is shown.

(ref. 22). Furthermore, only relatively low levels of the inhibitors may be required to produce a therapeutic effect given the acute sensitivity of the BRCA2-defective cells.

To test the hypothesis that PARP inhibitors may be useful for specific treatment of BRCA2-deficient tumours, we implanted BRCA2-deficient V-C8 and BRCA2-complemented V-C8+B2 cells intramuscularly in the thigh of 40 CD-1 nude mice. The xenograft uptake rate was 10 out of 20 for V-C8 cells and 9 out of 20 for V-C8+B2 cells. We monitored tumour growth by measuring the diameter of the implanted thigh compared with the diameter of the other non-implanted thigh. Saline (10 mg ml⁻¹) or two doses of AG14361 (25 or 50 mg kg⁻¹) in saline were administered intraperitoneally daily for 5 days to mice bearing V-C8 or V-C8+B2 tumours. We found that AG14361 had no effect on the outgrowth of the V-C8+B2 xenografts as compared to xenografts in the saline-treated mice (Table 1). In contrast, three out of five V-C8 xenografts responded to the 5-day AG14361 treatment and one mouse showed complete remission with no sign of tumour at autopsy (Fig. 4). Although not all V-C8 xenografts responded to treatment, these experiments clearly show that BRCA2-deficient tumours in a xenograft model are susceptible to treatment with the PARP inhibitor AG14361 alone. Our data provide one of the few examples where the defect causing the cancer is exploited for targeted therapy. Furthermore, we provide the first example of a case where inhibitors of a normally non-essential DNA repair protein can be used alone to treat a tumour, which is a new concept in cancer treatment. □

Methods

Cell culture

Cell lines used were MCF-7 and MDA-MB-231 (American Type Culture Collection), and irs1 (ref. 10), irs1X2.1 (ref. 10), V79-4 (ref. 10), AA8 (ref. 11), EM9 (ref. 21), irs1SF (ref. 11), CXR3 (ref. 11), V-C8 (ref. 14), V79 (ref. 14), V-C8 + B2 (ref. 14), V-C8#13 (ref. 14) and SW480SN.3 (ref. 23). All cells were grown in Dulbecco's modified Eagle's Medium (DMEM) with 10% fetal bovine serum, penicillin (100 U ml⁻¹) and streptomycin sulphate (100 µg ml⁻¹) at 37 °C under an atmosphere containing 5% CO₂.

Immunofluorescence

Cells were plated onto coverslips 24 h before 24-h treatments, as indicated. After treatments, the medium was removed, coverslips rinsed once in PBS at 37 °C and fixed and stained as described previously⁸.

The primary antibodies used in this study were rabbit polyclonal anti-γ-H2AX (Trevigen) at a dilution of 1:500, and rabbit polyclonal anti-RAD51 (H-92, Santa Cruz) at a dilution of 1:1,000. The secondary antibody was Alexa 555 goat anti-rabbit F(ab')₂ IgG or Alexa 555 donkey anti-rabbit IgG antibody (Molecular Probes), both at a concentration of 1:500. Antibodies were diluted in PBS containing 3% bovine serum albumin. DNA was stained with 1 µg ml⁻¹ To-Pro (Molecular Probes).

Images were obtained with a Zeiss LSM 510 inverted confocal microscope using a planapochromat × 63/NA 1.4 oil immersion objective and excitation wavelengths of 546 and 630 nm. Through focus maximum projection images were acquired from optical sections 0.50 µm apart and with a section thickness of 1.0 µm. Images were processed using Adobe PhotoShop (Abacus Inc).

The frequencies of cells containing γ-H2AX or RAD51 foci were determined in at least three separate experiments. At least 300 nuclei were counted on each slide. Nuclei containing more than five γ-H2AX foci or ten RAD51 foci were classified as positive.

PARP1 activity assays

PARP1 activity was assayed in permeabilized cells using a modification of the method in ref. 24. We determined PARP1 inhibition by NU1025 and AG14361 in intact log-phase cells before digitonin permeabilization and stimulation of PARP1 activity by a short palindromic oligonucleotide (CGGAATCCG)²⁵ in the presence of 75 µM [³²P]NAD⁺, as described previously²².

siRNA treatment

Pre-designed BRCA2 SMARTpool and scrambled siRNAs were purchased from Dharmacon, and PARP1 and PARP2 siRNAs were designed and purchased from Dharmacon. 1.5 × 10⁵ or 2.5 × 10⁴ cells were seeded onto 6- or 24-well plates, respectively, and incubated overnight before transfection with 200 nM siRNA using oligofectamine (Invitrogen) according to the manufacturer's instructions. Cells were then cultured in normal growth media for 48 h before trypsinization and re-plating for toxicity or colony outgrowth assays. Suppression of BRCA2 was confirmed by western blotting (as described previously²⁶) of protein extracts treated with siRNA with an antibody against BRCA2 (Oncogene) or PARP1 (Santa Cruz Biotech), or by RT-PCR analysis (Abgene) using PARP1 forward primer 5'-ATATTCCTCCGAGAACACAGC-3', PARP1 reverse

primer 5'-CTAATCTTTTGTAGCCCCCTT-3', PARP2 forward primer 5'-GTGGCAGAAATCTT GGCTGGT-3', PARP2 reverse primer 5'-GTGGTCTAAGGCGTACTGAA-3', BRCA2 forward primer 5'-TGATCCAAAGGGTCCCAAAGTTTC-3', BRCA2 reverse primer 5'-TTTACAGCTTTTTCAGAGCCTCACA-3', β-actin forward primer 5'-ACA CCCCAGCATGTAGCTAGGCC-3', and β-actin reverse primer 5'-AAGAGCCTCAGGG CAACGGAACC-3'. RT-PCR conditions were 42 °C, 30 min for one cycle; 95 °C, 5 min for one cycle; 95 °C, 30 s, 55 °C (BRCA2) or 60 °C (PARP1,2 and β-actin) for 20 s, and 72 °C, 20 s for variable cycle numbers so no band was saturated; and 72 °C, 5 min for one cycle.

Cytotoxicity assay

A total of 500 cells were plated in triplicate onto 100-mm dishes or 200 cells were plated in triplicate into 6-well plates 4 h before the addition of PARP inhibitors, as indicated. Treatment was for 24 h, after which cells were washed in PBS and fresh media added, or continuous, in which case original media containing the inhibitor was left on cells until colonies were counted. In treatments with AG14361, exponentially growing cells were treated for 24 h, trypsinized and re-seeded onto Petri dishes. Colonies were fixed and stained with methylene blue in methanol (4 g l⁻¹) or crystal violet in methanol:acetic acid 3:1 (4 g l⁻¹) 7–12 days later. Colonies consisting of more than 50 cells were subsequently counted.

Xenograft studies

V-C8 or V-C8 + B2 cells were implanted intramuscularly into the thigh of 40 CD-1 nude mice. Treatments were initiated when tumours were of measurable size (approximate leg diameter of 11 mm). Animals received either AG14361 or saline intraperitoneally administered on days 1–5, and were monitored on a daily basis during treatment (tumour measurements, body weights and clinical evidence recorded) and as required after the last treatment.

Received 14 October 2004; accepted 9 February 2005; doi:10.1038/nature03443.

- Lindahl, T., Satoh, M. S., Poirier, G. G. & Klungland, A. Post-translational modification of poly(ADP-ribose) polymerase induced by DNA strand breaks. *Trends Biochem. Sci.* **20**, 405–411 (1995).
- D'Amours, D., Desnoyers, S., D'Silva, I. & Poirier, G. G. Poly(ADP-ribosylation) reactions in the regulation of nuclear functions. *Biochem. J.* **342**, 249–268 (1999).
- El-Khamisy, S. F., Masutani, M., Suzuki, H. & Caldecott, K. W. A requirement for PARP-1 for the assembly or stability of XRCC1 nuclear foci at sites of oxidative DNA damage. *Nucleic Acids Res.* **31**, 5526–5533 (2003).
- Conde, C. *et al.* Loss of poly(ADP-ribose) polymerase-1 causes increased tumour latency in p53-deficient mice. *EMBO J.* **20**, 3535–3543 (2001).
- de Murcia, J. M. *et al.* Requirement of poly(ADP-ribose) polymerase in recovery from DNA damage in mice and in cells. *Proc. Natl Acad. Sci. USA* **94**, 7303–7307 (1997).
- Wang, Z. Q. *et al.* PARP is important for genomic stability but dispensable in apoptosis. *Genes Dev.* **11**, 2347–2358 (1997).
- Molinete, M. *et al.* Overproduction of the poly(ADP-ribose) polymerase DNA-binding domain blocks alkylation-induced DNA repair synthesis in mammalian cells. *EMBO J.* **12**, 2109–2117 (1993).
- Schultz, N., Lopez, E., Saleh-Gohari, N. & Helleday, T. Poly(ADP-ribose) polymerase (PARP-1) has a controlling role in homologous recombination. *Nucleic Acids Res.* **31**, 4959–4964 (2003).
- Yang, Y. G., Cortes, U., Patnaik, S., Jasin, M. & Wang, Z. Q. Ablation of PARP-1 does not interfere with the repair of DNA double-strand breaks, but compromises the reactivation of stalled replication forks. *Oncogene* **23**, 3872–3882 (2004).
- Griffin, C. S., Simpson, P. J., Wilson, C. R. & Thacker, J. Mammalian recombination-repair genes XRCC2 and XRCC3 promote correct chromosome segregation. *Nature Cell Biol.* **2**, 757–761 (2000).
- Tebbs, R. S. *et al.* Correction of chromosomal instability and sensitivity to diverse mutagens by a cloned cDNA of the XRCC3 DNA repair gene. *Proc. Natl Acad. Sci. USA* **92**, 6354–6358 (1995).
- Bowman, K. J., Newell, D. R., Calvert, A. H. & Curtin, N. J. Differential effects of the poly(ADP-ribose) polymerase (PARP) inhibitor NU1025 on topoisomerase I and II inhibitor cytotoxicity in L1210 cells *in vitro*. *Br. J. Cancer* **84**, 106–112 (2001).
- Skalitzky, D. J. *et al.* Tricyclic benzimidazoles as potent poly(ADP-ribose) polymerase-1 inhibitors. *J. Med. Chem.* **46**, 210–213 (2003).
- Kraakman-van der Zwet, M. *et al.* Brca2 (XRCC11) deficiency results in radioresistant DNA synthesis and a higher frequency of spontaneous deletions. *Mol. Cell Biol.* **22**, 669–679 (2002).
- Farmer, H. *et al.* Targeting the DNA repair defect in BRCA mutant cells as a therapeutic strategy. *Nature* doi:10.1038/nature03445 (this issue).
- Venkitaraman, A. R. Cancer susceptibility and the functions of BRCA1 and BRCA2. *Cell* **108**, 171–182 (2002).
- Menissier de Murcia, J. *et al.* Functional interaction between PARP-1 and PARP-2 in chromosome stability and embryonic development in mouse. *EMBO J.* **22**, 2255–2263 (2003).
- Lindahl, T. Instability and decay of the primary structure of DNA. *Nature* **362**, 709–715 (1993).
- Paull, T. T. *et al.* A critical role for histone H2AX in recruitment of repair factors to nuclear foci after DNA damage. *Curr. Biol.* **10**, 886–895 (2000).
- Arnaudeau, C., Lundin, C. & Helleday, T. DNA double-strand breaks associated with replication forks are predominantly repaired by homologous recombination involving an exchange mechanism in mammalian cells. *J. Mol. Biol.* **307**, 1235–1245 (2001).
- Caldecott, K. W., Tucker, J. D. & Thompson, L. H. Construction of human XRCC1 minigenes that fully correct the CHO DNA repair mutant EM9. *Nucleic Acids Res.* **20**, 4575–4579 (1992).
- Calabrese, C. R. *et al.* Anticancer chemosensitization and radiosensitization by the novel poly(ADP-ribose) polymerase-1 inhibitor AG14361. *J. Natl Cancer Inst.* **96**, 56–67 (2004).
- Mohindra, A. *et al.* Defects in homologous recombination repair in mismatch-repair-deficient tumour cell lines. *Hum. Mol. Genet.* **11**, 2189–2200 (2002).
- Halldorsson, H., Gray, D. A. & Shall, S. Poly (ADP-ribose) polymerase activity in nucleotide permeable cells. *FEBS Lett.* **85**, 349–352 (1978).
- Grube, K., Kupper, J. H. & Burkle, A. Direct stimulation of poly(ADP ribose) polymerase in

permeabilized cells by double-stranded DNA oligomers. *Anal. Biochem.* **193**, 236–239 (1991).
26. Lundin, C. *et al.* RAD51 is involved in repair of damage associated with DNA replication in mammalian cells. *J. Mol. Biol.* **328**, 521–535 (2003).

Supplementary Information accompanies the paper on www.nature.com/nature.

Acknowledgements We wish to thank J. Lunec, J. Thacker, L. Thompson, M. Zdzienicka, Z. Hostomsky and Pfizer GRD, La Jolla for providing materials. The investigation was financed by grants to T.H. and M.M. from Yorkshire Cancer Research. Additional support was financed through grants to T.H. from the Swedish Cancer Society and the Swedish Research Council and a grant to N.J.C. from Cancer Research-UK.

Competing interests statement The authors declare that they have no competing financial interests.

Correspondence and requests for materials should be addressed to T.H. (t.helleday@sheffield.ac.uk).

Targeting the DNA repair defect in *BRCA* mutant cells as a therapeutic strategy

Hannah Farmer^{1,2*}, Nuala McCabe^{1,2*}, Christopher J. Lord^{2*}, Andrew N. J. Tutt^{2,3}, Damian A. Johnson², Tobias B. Richardson², Manuela Santarosa^{2†}, Krystyna J. Dillon⁴, Ian Hickson⁴, Charlotte Knights⁴, Niall M. B. Martin⁴, Stephen P. Jackson^{4,5}, Graeme C. M. Smith⁴ & Alan Ashworth^{1,2}

¹Cancer Research UK Gene Function and Regulation Group and ²The Breakthrough Breast Cancer Research Centre Institute of Cancer Research, Fulham Road, London SW3 6JB, UK

³Guy's Hospital, St Thomas' Street, London SE1 9RT, UK

⁴KuDOS Pharmaceuticals Ltd, Cambridge Science Park, Cambridge CB4 0WG, UK

⁵Wellcome Trust and Cancer Research UK, Gurdon Institute of Cancer and Developmental Biology, and Department of Zoology, University of Cambridge, Tennis Court Road, Cambridge CB2 1QN, UK

* These authors contributed equally to this work

† Present address: Division of Experimental Oncology1, CRO-IRCCS, Aviano 33081 PN, Italy

BRCA1 and BRCA2 are important for DNA double-strand break repair by homologous recombination¹, and mutations in these genes predispose to breast and other cancers². Poly(ADP-ribose) polymerase (PARP) is an enzyme involved in base excision repair, a key pathway in the repair of DNA single-strand breaks³. We show here that BRCA1 or BRCA2 dysfunction unexpectedly and profoundly sensitizes cells to the inhibition of PARP enzymatic activity, resulting in chromosomal instability, cell cycle arrest and subsequent apoptosis. This seems to be because the inhibition of PARP leads to the persistence of DNA lesions normally repaired by homologous recombination. These results illustrate how different pathways cooperate to repair damage, and suggest that the targeted inhibition of particular DNA repair pathways may allow the design of specific and less toxic therapies for cancer.

Loss of PARP1 function can induce the formation of nuclear RAD51 foci as well as sister chromatid exchanges without increasing homologous recombination⁴. This suggests that loss of PARP1 increases the formation of DNA lesions that are repaired by homologous recombination without directly regulating the process of homologous recombination itself. As BRCA1 and BRCA2 are important for homologous recombination^{5–7}, we hypothesized that inhibition of PARP1 in a BRCA1- or BRCA2-defective background might result in the generation of DNA lesions normally repaired by

sister chromatid exchange, causing chromatid aberrations and loss of viability. To examine the effects of PARP1 depletion, we transfected a plasmid expressing a short interfering RNA (siRNA) targeting mouse *Parp1* into wild-type embryonic stem (ES) cells and ES cells lacking wild-type *Brca1* or *Brca2* (refs 7, 8). These cells bear specific genomic mutations of *Brca1* or *Brca2* and like *BRCA1/2* tumours lack a wild-type allele and can be directly compared to their isogenic wild-type counterparts. The *Parp1* siRNA construct caused a clear reduction in clonogenic survival of BRCA1- and BRCA2-deficient cells compared with wild-type cells (Fig. 1a, b). This prompted us to test whether chemical inhibitors of PARP activity might have similar effects. We used two novel, specific and very potent small-molecule PARP inhibitors: KU0058684 (PARP1 half-maximal inhibitory concentration (IC₅₀) = 3.2 nM) and KU0058948 (PARP1 IC₅₀ = 3.4 nM), as well as a much less active but chemically related compound KU0051529 (PARP1 IC₅₀ = 730 nM) (Supplementary Fig. 1 and Supplementary Table 1).

Clonogenic cell survival assays showed that BRCA1- or BRCA2-deficient cells were extremely sensitive to KU0058684 and KU0058948 compared with heterozygous mutant or wild-type cells (Fig. 1c), and that these effects were rapid and irreversible (Supplementary Fig. 2). The SF₅₀ (dosage at which 50% of cells survived) for KU0058684 was 35 nM for BRCA1-deficient ES cells and 15 nM for BRCA2-deficient ES cells; for wild-type cells this was approximately 2 μ M. Notably, this represents factors of 57-fold and 133-fold enhanced sensitivity of cells lacking wild-type BRCA1 and BRCA2, respectively, compared with wild-type-cells. Similar results were obtained with non-embryonic cells such as BRCA2-deficient Chinese hamster ovary cells⁹, which showed a greater than 1,000-fold enhanced sensitivity compared with a BRCA2-complemented derivative (Supplementary Fig. 3a). Similarly, depletion of *BRCA1* messenger RNA in MCF7 human breast cancer cells by RNA interference induced sensitivity to PARP inhibition (Supplementary Fig. 3b). In contrast, KU0051529, which does not effectively inhibit PARP1 or PARP2, had no selective effect on cells lacking wild-type BRCA1 or BRCA2. In conjunction with the siRNA data, this indicates that the mechanism of sensitivity is through inhibition of PARP. Although BRCA1 and BRCA2 mutant cells do show enhanced sensitivity to certain DNA-damaging agents, such as cisplatin, this was to a much lesser degree (approximately threefold) (data not shown), demonstrating the greater selectivity of PARP inhibition compared to chemotherapy. Notably, none of the inhibitors had any selective effect on cells heterozygous for *Brca1* or *Brca2* mutations (Fig. 1c).

After 24 h exposure, KU0058684 elicited a profound arrest of cells with a tetraploid DNA content, indicating arrest in the G2 or M phase of the cell cycle (Fig. 1d). Most (>85%) of the arrested cells did not become labelled with an anti-phospho histone H3 antibody, an M-phase marker, indicating predominant arrest at the G2 stage (data not shown). Fluorescence-activated cell sorting (FACS) analysis of annexin-V-stained ES cells after 48 h treatment with KU0058684 showed a considerable increase in apoptosis (Fig. 1e). Together, these data suggest that cell cycle arrest followed by apoptosis leads to the reduced cell survival observed in the clonogenic assays. Examination of mitotic chromosomes of ES cells lacking wild-type BRCA1 or BRCA2 revealed that 10 nM or 1 μ M KU0058684 treatment resulted in frequent major aberrations (Fig. 1f). These included chromatid breaks and more complex chromatid aberrations such as tri-radial and quadri-radial chromosomes. At the same dose, such aberrations were rarely seen in wild-type cells (Supplementary Fig. 4a). These phenotypes are suggestive of a failure to repair double-strand breaks (DSBs) by conservative RAD51-dependent sister chromatid recombination and the consequent use of alternative error-prone pathways such as single-strand annealing and/or non-homologous end joining.

Formation of nuclear γ -H2AX foci occurs at sites of DNA damage, including DSBs and arrested replication forks¹⁰. Treatment with KU0058684 induces the formation of γ -H2AX foci (Fig. 2a; see also Supplementary Fig. 4b, c) to the same extent in wild-type and BRCA1- or BRCA2-deficient cells, suggesting that the damage induced is independent of BRCA function. Importantly, γ -H2AX

foci formation also occurs after RNA interference-mediated silencing of PARP1 (Supplementary Fig. 4d), indicating that KU0058684 is unlikely to act through a dominant-negative effect. To determine how these lesions might be repaired, we investigated whether KU0058684 elicited RAD51 focus formation. The formation of these foci is one hallmark of BRCA- dependent DSB repair, and

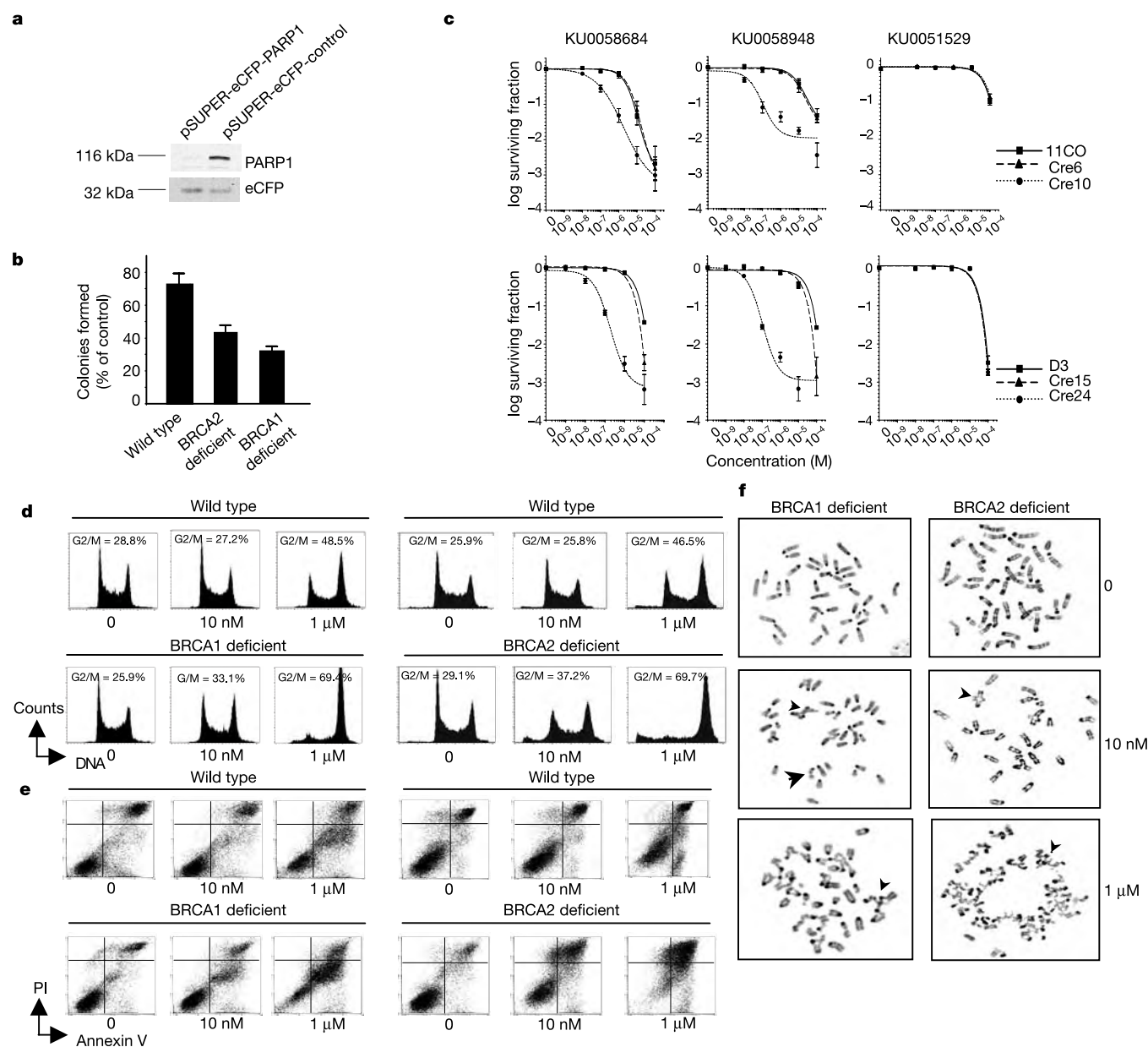


Figure 1 Depletion or inhibition of PARP1 selectively reduces the viability of BRCA1- and BRCA2-deficient ES cells. **a**, Reduction of PARP1 protein levels by siRNA. A plasmid (pSUPER-eCFP-PARP1) expressing a *Parp1*-specific siRNA and eCFP was transfected into ES cells. A control plasmid (pSUPER-eCFP-control), expressing an unrelated scrambled siRNA and eCFP, was separately transfected. Cell lysates were analysed by western blotting with either an anti-PARP1 antibody or anti-CFP antiserum as a control. **b**, Reduction in viability of BRCA1- and BRCA2-deficient ES cells after *Parp1*-specific siRNA silencing. ES cells were transfected with either pSUPER-eCFP-PARP1 or pSUPER-eCFP-control together with a blasticidin-resistance-encoding plasmid, and resistant clones counted. Error bars represent one standard deviation around the mean. Wild-type cells are significantly less sensitive than BRCA1- or BRCA2-deficient cells (Student's *t*-test $P < 0.05$). **c**, Inhibition of PARP activity selectively inhibits the survival of cells lacking wild-type BRCA1 or BRCA2. Clonogenic survival curves of BRCA1 wild-type

(11CO), heterozygous (Cre6) and deficient (Cre10) ES cells, and BRCA2 wild-type (D3), heterozygous (Cre15) and deficient (Cre24) ES cells after 10–12 days continuous exposure to chemical inhibitors. Error bars represent one standard deviation around the mean. **d**, Exposure to PARP inhibitor for 24 h results in G2/M arrest. ES cells were treated with 10 nM or 1 μ M KU0058684 or vehicle and analysed by FACS. **e**, Exposure to PARP inhibitor for 48 h results in apoptosis. ES cells were treated with KU0058684 or vehicle and stained with an annexin V antibody and propidium iodide (PI). The lower right quadrant of each panel represents early apoptotic cells and the upper right quadrant represents late apoptotic/necrotic cells. **f**, Chromosome analysis of BRCA1- and BRCA2-deficient ES cells exposed to KU0058684 or vehicle for 24 h. The small arrowheads indicate complex chromatid rearrangements and the large arrowhead indicates a chromatid break.

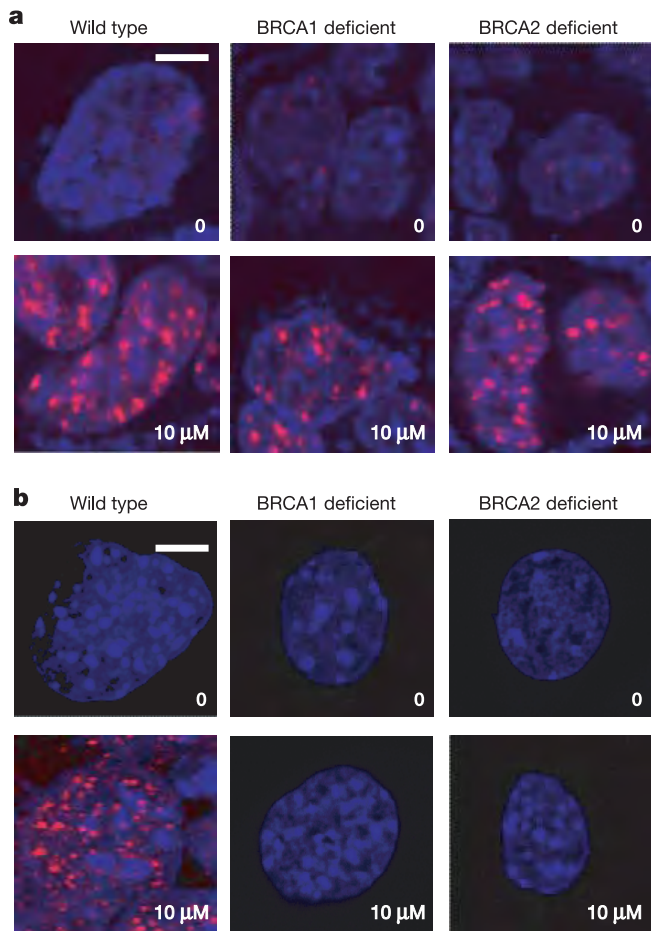


Figure 2 DSB formation and repair after exposure to PARP inhibitor. **a**, ES cells were exposed to 10 μ M KU0058684 or vehicle for 48 h. Nuclei are shown in blue and γ -H2AX foci in red. **b**, ES cells were exposed to 10 μ M KU0058684 or vehicle for 48 h. Nuclei are shown in blue and RAD51 foci in red. Scale bar, 10 μ m.

the loss of PARP1 function is known to induce the appearance of these foci⁴. In wild-type ES cells, KU0058684 caused formation of RAD51 foci in a dose-dependent fashion (Fig. 2b; see also Supplementary Fig. 4e). In contrast, no foci were formed in cells lacking wild-type BRCA1 or BRCA2 regardless of drug dose. This latter finding is consistent with previous observations that agents that cause recombinogenic lesions do not induce formation of RAD51 foci in BRCA1- or BRCA2-deficient cells^{7,11}. It seems likely that KU0058684 induces DNA damage normally repaired by a mechanism that involves RAD51 and that requires BRCA1 and BRCA2.

We tested the *in vivo* efficacy of our PARP inhibitor in preventing the formation of BRCA2-deficient tumours. Existing tumour cell models of BRCA1 and BRCA2 deficiency are unsuitable for xenograft growth for the testing of small-molecule therapeutics. Therefore, we used the ability of ES cells to form teratocarcinomas after transplantation into athymic mice. KU0058684 treatment severely inhibited the formation of tumours derived from BRCA2-deficient cells (Fig. 3). This inhibition was selective, as animals injected with wild-type cells and treated with KU0058684 continued to succumb to tumours. This demonstrates that the selective activity of PARP inhibitors on BRCA2-deficient cells observed *in vitro* can be reproduced *in vivo*, arguing for the therapeutic potential of this approach.

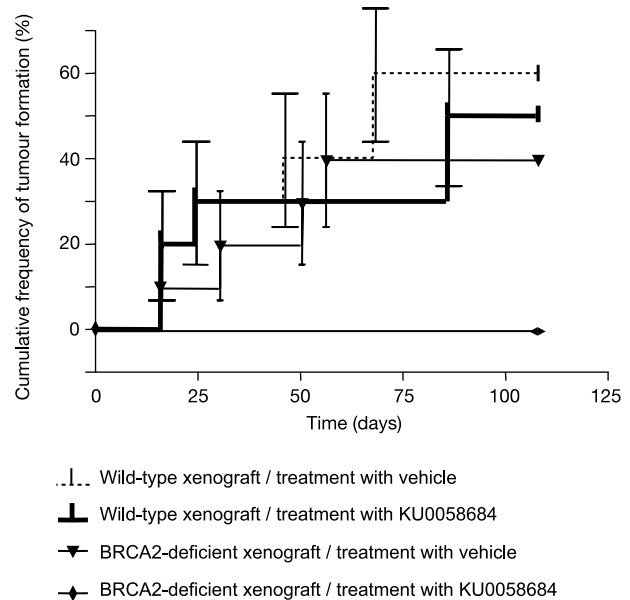


Figure 3 PARP inhibition selectively blocks the growth of BRCA2-deficient tumours *in vivo*. BRCA2-deficient ES cells or wild-type cells were injected into nude mice. Animals were then treated with either KU0058684 or vehicle. Significant differences in tumour formation were seen between the BRCA2-deficient xenograft/KU0058684 treatment cohort and the BRCA2-deficient xenograft/vehicle treatment cohort ($P = 0.03$) and also between the BRCA2-deficient xenograft/KU0058684 treatment cohort and the wild-type xenograft/KU0058684 treatment cohort ($P = 0.01$). Similar data were obtained in an independent experiment. Error bars represent one standard deviation around the mean.

Why might BRCA1- and BRCA2-deficient cells exhibit extreme sensitivity to PARP inhibition? We propose a model (Fig. 4) that arises from the observations that PARP is required for the efficient repair of DNA single-strand breaks (SSBs) during base excision repair^{3,12} and that PARP inhibition leads to persistent single-strand gaps in DNA¹³. If these gaps are encountered by a replication fork, arrest would occur and the single-strand gaps may degenerate into DSBs¹⁴. Normally these DSBs can be repaired by RAD51-dependent homologous recombination¹⁵, a process in which both BRCA1 and BRCA2 are involved^{1,11}. In the absence of BRCA1 or BRCA2, the replication fork cannot be restarted and collapses¹⁶, causing persistent chromatid breaks. Repair of these breaks by alternative error-prone DSB repair mechanisms would cause large numbers of chromatid breaks and aberrations, leading to loss of viability (Fig. 4).

The results presented here and in the accompanying paper¹⁷ suggest a new mechanism-based approach for the treatment of patients with BRCA1- and BRCA2-associated cancers. Tumours in carriers of BRCA1 or BRCA2 mutations lack wild-type BRCA1 or BRCA2 but normal tissues retain a single wild-type copy of the relevant gene. This biochemical difference provides the rationale for an approach involving PARP inhibition to generate specific DNA lesions that require functional BRCA1 and BRCA2 for their repair, generating a potentially large therapeutic window. This approach is likely to be less toxic and more specific than standard cytotoxic chemotherapy. PARP inhibitors are relatively non-toxic and do not directly damage DNA, and *Parp1* knockout mice are viable and healthy¹⁸. Our model for sensitivity to PARP inhibition depends on homologous recombination deficiency and not on inherited BRCA1 or BRCA2 deficiency *per se*. Therefore, this approach may be more widely applicable in the treatment of sporadic cancers with

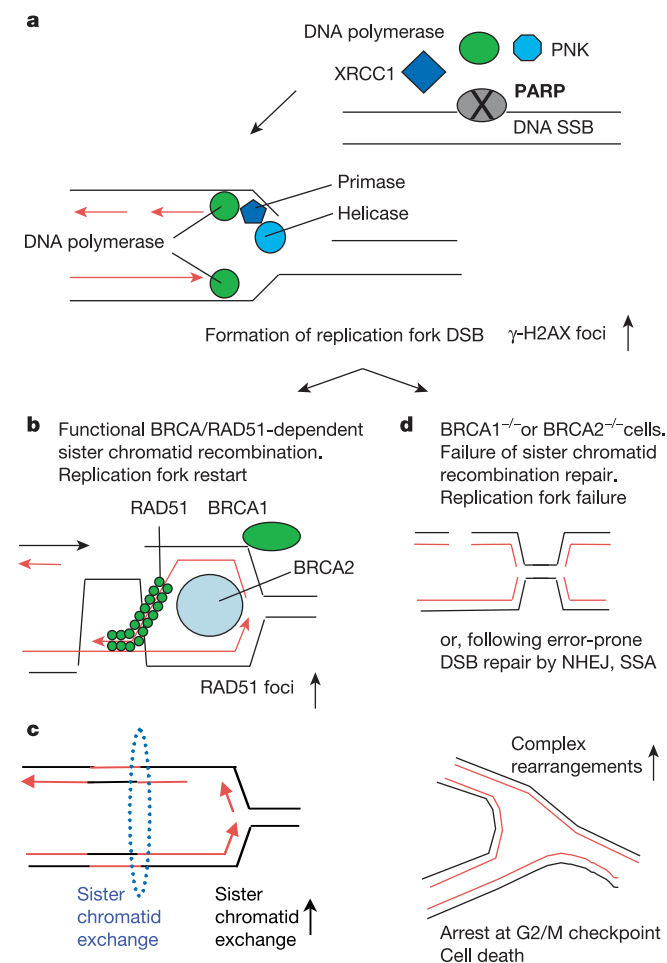


Figure 4 A model for the selective effects of PARP inhibition on cells lacking wild-type BRCA1 and BRCA2. **a**, PARP functions in base excision repair. DNA SSBs form due to oxidative damage and its repair. Inhibition of PARP activity prevents the recruitment of XRCC1 and subsequent SSB gap filling by DNA polymerases. Large numbers of DNA SSBs persist and are encountered by DNA replication forks. These lead to replication fork arrest associated with a DSB. **b**, In the presence of functional BRCA1 and BRCA2, the proximity of an undamaged sister chromatid template to the DSB allows the invasion of the sister chromatid by a RAD51-coated single-stranded DNA filament, and initiation of sister chromatid recombination repair. This is associated with the formation of nuclear RAD51 foci. A collapsed replication fork may be restarted by this mechanism. **c**, When Holliday junctions at recombination intermediates are resolved, a sister chromatid exchange may occur. The excess number of replication fork arrests associated with loss of PARP function leads to an increase in sister chromatid recombination events and sister chromatid exchanges. **d**, In the absence of functional BRCA1 or BRCA2, sister chromatid recombination and the formation of RAD51 foci are severely impaired. Replication-associated DSBs cannot be repaired by sister chromatid recombination. Some remain unrepaired as chromatid breaks but many are repaired by error-prone RAD51-independent mechanisms such as non-homologous end joining (NHEJ) and single-strand annealing (SSA). These cause complex chromatid rearrangements. These cells arrest at the G2/M checkpoint and permanently arrest or undergo apoptosis.

'BRCAness'¹⁹ or other impairments of the homologous recombination pathway. □

Methods

Cell lines

ES cells lacking wild-type BRCA2 have been described previously⁷. ES cells lacking wild-type BRCA1 will be described elsewhere (A. Gabriel and A.A., unpublished data) but have previously been validated⁸. They carry a homozygous deletion of exons 22–24 of *Brca1*.

Small-molecule inhibitors of PARP

Details of the synthesis of PARP inhibitors have been described²⁰. IC₅₀ values for the compounds *in vitro* were generated using assays previously described for PARP proteins^{21–23}.

Clonogenic assays

ES cells were maintained on tissue culture dishes coated with 0.1% gelatin and were transfected with either pSUPER-eCFP-PARP1 or pSUPER-eCFP-control along with a vector conferring resistance to blasticidin (pEF-Bsd, Invitrogen). Twenty-four hours after transfection, cells were seeded in 6-well plates. Forty-eight hours after transfection, treatment with blasticidin was started. After 10–14 days, cells were fixed in methanol and stained with crystal violet. Colonies were counted using a ColCount machine (Oxford Optronix).

For measurement of sensitivity to chemical inhibitors, exponentially growing cells were seeded at various densities in 6-well plates onto mitomycin C-inactivated mouse embryonic fibroblasts. Where appropriate, cells were treated with inhibitors after 18 h. For continuous exposure with inhibitor, cells were re-fed every 4 days with fresh medium and inhibitor. After 10–14 days, cells were fixed, stained with crystal violet and counted as above.

Xenografts

ES-cell-derived tumours (teratomas) were produced by subcutaneous injection of 2×10^6 ES cells into 6–8 week athymic BALB/c-nude (nu/nu) mice. Forty mice were injected with BRCA2-deficient ES cells or isogenic wild-type cells. Two days after cell injection, treatment with KU0058684 or vehicle was initiated. For three consecutive days, two intraperitoneal doses of KU0058684 (or vehicle) were administered, 6 h apart, each at a dosage of 15 mg kg^{-1} per animal. This treatment was then stopped for 5 days and then re-initiated (as before) for another three consecutive days. Growth of tumours was monitored from a minimum volume of 0.2 cm^3 .

Received 14 October 2004; accepted 11 February 2005; doi:10.1038/nature03445.

1. Tutt, A. & Ashworth, A. The relationship between the roles of *BRCA* genes in DNA repair and cancer predisposition. *Trends Mol. Med.* **8**, 571–576 (2002).
2. Wooster, R. & Weber, B. Breast and ovarian cancer. *N. Engl. J. Med.* **348**, 2339–2347 (2003).
3. Hoeijmakers, J. H. Genome maintenance mechanisms for preventing cancer. *Nature* **411**, 366–374 (2001).
4. Schultz, N., Lopez, E., Saleh-Gohari, N. & Helleday, T. Poly(ADP-ribose) polymerase (PARP-1) has a controlling role in homologous recombination. *Nucleic Acids Res.* **31**, 4959–4964 (2003).
5. Moynahan, M. E., Pierce, A. J. & Jasin, M. BRCA2 is required for homology-directed repair of chromosomal breaks. *Mol. Cell* **7**, 263–272 (2001).
6. Moynahan, M. E., Chiu, J. W., Koller, B. H. & Jasin, M. Brca1 controls homology-directed DNA repair. *Mol. Cell* **4**, 511–518 (1999).
7. Tutt, A. *et al.* Mutation in *Brca2* stimulates error-prone homology-directed repair of DNA double-strand breaks occurring between repeated sequences. *EMBO J.* **20**, 4704–4716 (2001).
8. Foray, N. *et al.* A subset of ATM- and ATR-dependent phosphorylation events requires the BRCA1 protein. *EMBO J.* **22**, 2860–2871 (2003).
9. Kraakman-van der Zwet, M. *et al.* *Brca2* (XRCC11) deficiency results in radioresistant DNA synthesis and a higher frequency of spontaneous deletions. *Mol. Cell Biol.* **22**, 669–679 (2002).
10. Furuta, T. *et al.* Phosphorylation of histone H2AX and activation of Mre11, Rad50, and Nbs1 in response to replication-dependent DNA double-strand breaks induced by mammalian DNA topoisomerase I cleavage complexes. *J. Biol. Chem.* **278**, 20303–20312 (2003).
11. Bhattacharyya, A., Ear, U. S., Koller, B. H., Weichselbaum, R. R. & Bishop, D. K. The breast cancer susceptibility gene *BRCA1* is required for subnuclear assembly of Rad51 and survival following treatment with the DNA cross-linking agent cisplatin. *J. Biol. Chem.* **275**, 23899–23903 (2000).
12. Dantzer, F. *et al.* Involvement of poly(ADP-ribose) polymerase in base excision repair. *Biochimie* **81**, 69–75 (1999).
13. Boulton, S., Kyle, S., & Durkacz, B. W. Interactive effects of inhibitors of poly(ADP-ribose) polymerase and DNA-dependent protein kinase on cellular DNA damage. *Carcinogenesis* **20**, 199–203 (1999).
14. Haber, J. E. DNA recombination: the replication connection. *Trends Biochem. Sci.* **24**, 271–275 (1999).
15. Arnaudeau, C., Lundin, C. & Helleday, T. DNA double-strand breaks associated with replication forks are predominantly repaired by homologous recombination involving an exchange mechanism in mammalian cells. *J. Mol. Biol.* **307**, 1235–1245 (2001).
16. Lomonosov, M., Anand, S., Sangrithi, M., Davies, R. & Venkitesan, A. R. Stabilization of stalled DNA replication forks by the BRCA2 breast cancer susceptibility protein. *Genes Dev.* **17**, 3017–3022 (2003).
17. Bryant, H. E. *et al.* Specific killing of BRCA2-deficient tumours with inhibitors of poly(ADP-ribose) polymerase. *Nature* doi:10.1038/nature03443 (this issue).
18. Wang, Z. Q. *et al.* PARP is important for genomic stability but dispensable in apoptosis. *Genes Dev.* **11**, 2347–2358 (1997).
19. Turner, N., Tutt, A. & Ashworth, A. Hallmarks of BRCAness in sporadic cancers. *Nature Rev. Cancer* **4**, 814–819 (2004).
20. Loh, V. M. *et al.* Phthalazinones. Part I: The design and synthesis of a novel series of potent inhibitors of poly(ADP-ribose) polymerase. *Bioinorg. Med. Chem. Lett.* (in the press).
21. Ame, J.-C. *et al.* PARP-2, a novel mammalian DNA damage-dependent poly(ADP-ribose) polymerase. *J. Biol. Chem.* **274**, 17860–17868 (1999).
22. Kickhoefer, V. A. *et al.* The 193-kD vault protein, VPAP, is a novel poly(ADP-ribose) polymerase. *J. Cell Biol.* **146**, 917–928 (1999).
23. Dillon, K. J., Smith, G. C. M. & Martin, N. M. B. A flashplate assay for the identification of PARP-1 inhibitors. *J. Biomol. Screen.* **8**, 347–352 (2003).

Supplementary Information accompanies the paper on www.nature.com/nature.

Acknowledgements We thank Cancer Research UK, Breakthrough Breast Cancer and the Mary-Jean Mitchell Green Foundation for financial support. We thank I. Titley for help with FACS analysis, A. McCarthy and J. Williamson for help with chromosome spreads, E. Witt for western blot analysis, E. Iorns for real-time PCR analysis and M. Zdzienicka for V-C8 and V-C8 BAC cells. We also thank the Maybridge Chemical Company for their help in the design and synthesis of the PARP inhibitors.

Competing interests statement The authors declare competing financial interests: details accompany the paper on www.nature.com/nature.

Correspondence and requests for materials should be addressed to A.A. (alan.ashworth@icr.ac.uk) or G.C.M.S. (gcmsmith@kudospharma.co.uk).

Transcriptional regulation of a metastasis suppressor gene by Tip60 and β -catenin complexes

Jung Hwa Kim^{1*}, Bogyou Kim^{1*}, Ling Cai^{2*}, Hee June Choi¹, Kenneth A. Ohgi², Chris Tran³, Charlie Chen³, Chin Ha Chung¹, Otmar Huber⁴, David W. Rose⁵, Charles L. Sawyers³, Michael G. Rosenfeld² & Sung Hee Baek¹

¹Department of Biological Sciences, College of Natural Sciences, Seoul National University, Seoul 151-742, South Korea

²Howard Hughes Medical Institute, University of California, San Diego, Department and School of Medicine, La Jolla, California 92093-0648, USA

³Howard Hughes Medical Institute, Department of Medicine, University of California, Los Angeles, Los Angeles, California 90095, USA

⁴Institute of Clinical Chemistry and Pathobiochemistry, Charité - Campus Benjamin Franklin, Hindenburgdamm 30, 12200 Berlin, Germany

⁵Department of Medicine, Division of Endocrinology and Metabolism, School of Medicine, University of California, San Diego, San Diego, California 92093, USA

* These authors contributed equally to this work

Defining the molecular strategies that integrate diverse signalling pathways in the expression of specific gene programmes that are critical in homeostasis and disease remains a central issue in biology. This is particularly pertinent in cancer biology because downregulation of tumour metastasis suppressor genes is a common occurrence^{1,2}, and the underlying molecular mechanisms are not well established. Here we report that the downregulation of a metastasis suppressor gene, *KAI1*, in prostate cancer cells involves the inhibitory actions of β -catenin, along with a reptin chromatin remodelling complex. This inhibitory function of β -catenin–reptin requires both increased β -catenin expression and recruitment of histone deacetylase activity. The coordinated actions of β -catenin–reptin components that mediate the repressive state serve to antagonize a Tip60 coactivator complex^{3–8} that is required for activation; the balance of these opposing complexes controls the expression of *KAI1* and metastatic potential. The molecular mechanisms underlying the antagonistic regulation of β -catenin–reptin and the Tip60 coactivator complexes for the metastasis suppressor gene, *KAI1*, are likely to be prototypic of a selective downregulation strategy for many genes, including a subset of NF- κ B target genes.

KAI1 is a member of the tetraspanin family, which is capable of inhibiting the progression of tumour metastasis without affecting primary tumorigenicity *in vivo*^{9,10}. Analysis of *KAI1* messenger RNA by reverse-transcriptase-mediated polymerase chain reaction (RT–PCR) revealed that its levels were increased by treatment with interleukin-1 β (IL-1 β) in normal prostate cells and tumorigenic prostate cells, but not in cell lines with highly metastatic behaviour from prostate, breast or colon (Fig. 1a). These data imply a possible tissue-selective role for *KAI1* in suppressing

metastasis of human prostate cancer, but the role seems not to involve either an allelic loss or a mutation, implying that the reduction of *KAI1* gene expression is at the transcriptional level¹¹.

We addressed the potential functional role of *KAI1* as a metastasis suppressor gene by metastatic assays *in vivo* by assessing the effect of restoring *KAI1* expression to LNCaP prostate cancer cells, which do not effectively express *KAI1* (Fig. 1a, b). The lung was dissected from each mouse and luciferase activity was recorded (Fig. 1b, c). Luciferase activities in prostate tissue were also recorded to verify that *KAI1*-expressing cells had not undergone apoptosis (Fig. 1d). Although the primary tumour weights as well as associated luciferase activities in prostate were comparable both in control and in *KAI1*-expressing cell tumours (Fig. 1d, and Supplementary Fig. S1), there was a significant decrease in lung metastases in the *KAI1*-expressing LNCaP cells. These results indicate that expression of *KAI1* in prostate cancer cells significantly suppresses the *in vivo* incidence of lung metastases.

To address the molecular mechanisms of differential expression of *KAI1* at the transcriptional level in non-metastatic and metastatic cancer cells, we performed a chromatin immunoprecipitation (ChIP) assay to monitor different cofactors recruited on the *KAI1* promoter (Fig. 2a). In RWPE1 cells, coincident with IL-1 β -dependent dismissal of the N-CoR/TAB2 co-repressor complex¹², the Tip60 coactivator was recruited and acetylated histones H3 and H4 were detected. However, in metastatic prostate cancer cells, the Tip60 coactivator complex was not recruited to the *KAI1* promoter, and the *KAI1* promoter remained inactive, even after release of the N-CoR co-repressor complex (Fig. 2a).

Examination of the levels of Tip60 expression indicated that both RNA and protein levels of Tip60 were downregulated in metastatic cancer cells (Fig. 2b, c). To investigate whether this difference in the level of Tip60 expression was sufficient to account for the failure of transcriptional activation of *KAI1*, we first overexpressed Tip60 in LNCaP cells (Supplementary Fig. S2a). Remarkably, with overexpression of Tip60 in the presence of IL-1 β , Tip60 was recruited on the *KAI1* promoter, and *KAI1* mRNA levels were upregulated (Fig. 2d, e).

Tip60 and β -catenin complexes possess both pontin and reptin chromatin remodelling complexes as common components^{3–8}; however, the roles of pontin and reptin in the Tip60 or β -catenin complexes have not been studied extensively. Pontin and reptin have been reported to function as antagonistic regulators of β -catenin signalling by using transient transfection and reporter gene assays^{3–7}. We therefore examined the possibility that reptin or pontin might be recruited together with Tip60 as a component of the functional regulation in both non-metastatic and metastatic cancer cells. As shown in Fig. 2f, in non-metastatic cancer cells both Tip60 and pontin, but not reptin, were recruited on the promoter. In contrast, in metastatic cancer cells reptin was recruited, as was β -catenin, but pontin was not recruited (Fig. 2f). A ChIP analysis performed in the metastatic cancer cells that stably overexpressed exogenous Tip60 coactivator revealed that Tip60 recruitment was now again observed on the *KAI1* promoter, with the pontin chromatin remodelling complex (Fig. 2f). The occupancy of the *KAI1* promoter by the Tip60 coactivator complex seemed to be mutually exclusive with an actively repressing β -catenin–reptin complex.

Immunoblot analysis (Fig. 2g) revealed that β -catenin expression was increased, whereas Tip60 exhibited a low level of expression in metastatic cancer cells. To determine whether *KAI1* downregulation in metastatic cancer cells is correlated with Tip60 downregulation and whether Tip60 and/or pontin are crucial for the transcriptional activation of *KAI1*, we silenced the expression of Tip60 or pontin in non-metastatic cells by using short hairpin RNA molecules (shRNA)^{13,14}. We verified specific knock-down effects of Tip60 and pontin by immunoblotting (Supplementary Fig. S2b, c). An shRNA against Tip60 resulted in inhibition of transcriptional activity of

KAI1 by IL-1 β , whereas knockdown of pontin did not (Fig. 2h, i).

To determine whether β -catenin–reptin is required for the down-regulation of *KAI1*, we overexpressed a constitutive active mutant of β -catenin in non-metastatic cells (Supplementary Fig. S3a)^{15,16}. *KAI1* induction was inhibited in β -catenin-transfected non-metastatic cells but not in mock-transfected cells after treatment with IL-1 β (Fig. 3a). In β -catenin-overexpressing cells, the quantity of Tip60 transcript was reduced and the Tip60 coactivator was no longer recruited on the *KAI1* promoter (Fig. 3b, c); instead, β -catenin and reptin were recruited on the promoter, correlated with repression (Fig. 3c). This finding supports the conclusion that high levels of nuclear β -catenin are sufficient to mediate down-regulation of *KAI1*, on the basis of an altered balance of these opposing complexes, in part because of downregulation of the Tip60 coactivator itself.

Increasing Tip60 expression restored *KAI1* expression (Fig. 2d), whereas increasing β -catenin expression inhibited *KAI1* expression (Fig. 3a). To prove that β -catenin can actively displace Tip60, causing repression of the *KAI1* promoter, we used RWPE1 cells stably overexpressing β -catenin to examine the effects of overexpressing Tip60 with the use of two-step ChIP assays (Fig. 3d),

finding that β -catenin–reptin was recruited only on the *KAI1* promoters that were Tip60-negative. These data support the suggestion that β -catenin–reptin actively represses the *KAI1* promoter by displacing Tip60 coactivator complex.

β -Catenin has a crucial role in development and homeostasis, and deregulated expression of β -catenin is involved in oncogenesis and tumour progression^{17–20}. β -Catenin has been shown to form a complex with NF- κ B, and repressed NF- κ B activity is often observed in human cancer cells²¹. Because *KAI1* has been reported to be an NF- κ B target gene²², we examined whether all NF- κ B target genes are repressed or whether a subset of NF- κ B target genes are repressed by the increased expression of β -catenin. To address this question, we tested two other NF- κ B target promoters, *Fas* and *intracellular adhesion molecule-1* (*ICAM-1*) by ChIP assay (Fig. 3e, f).

Whereas effects on the *Fas* promoter were similar to that observed on the *KAI1* promoter (Fig. 3e), the *ICAM-1* promoter was not affected by the expression of β -catenin, revealing distinct and gene-selective molecular strategies for the downregulation of NF- κ B target genes by β -catenin (Fig. 3f). Although Tip60 was not recruited on the *KAI1* and *Fas* promoters in the presence of high levels of β -catenin expression, levels of Tip60 expression

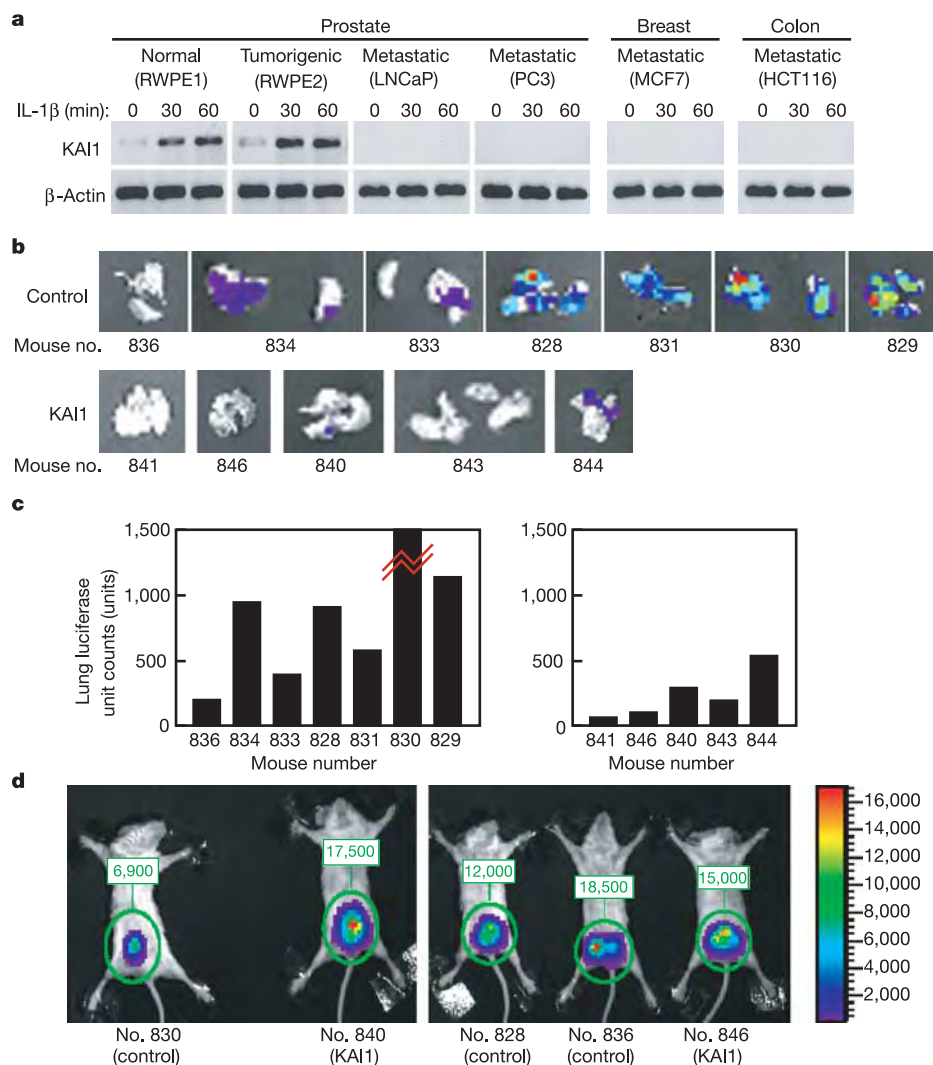


Figure 1 *KAI1* is a functional metastasis suppressor gene. **a**, *KAI1* message is downregulated in metastatic cancer cells. **b**, LNCaP cells expressing luciferase reporter with either empty vector (pLNCX2) or *KAI1* expression vector (*KAI1*-pLNCX2) were injected into mouse prostate, and luciferase activity in the lung, representative of lung metastasis, was recorded for each mouse. Lung images from different mice are shown.

c, Quantification of luciferase activities from either each control mouse (left) or each *KAI1*-expressing mouse (right). **d**, A control for luciferase activity still associated with the prostate is shown from each mouse. The colour bar represents luciferase intensity. Numbers relating to the green ovals represent luciferase unit counts in the prostate.

remained sufficient for the activation of other target gene promoters (Supplementary Fig. S3b).

To examine whether NF- κ B p50 is the DNA binding protein that is tethering Tip60 and the β -catenin–reptin complex to the *KAI1* promoter, we examined binding by ChIP to a thymidine kinase (tk) minimal promoter containing the three copies of the p50 response element present on the *KAI1* promoter (Fig. 3g, h). We found that, dependent on the p50 DNA-binding sites, p50 was recruited along with Bcl3, Tip60, β -catenin and reptin in the presence of IL-1 β (Fig. 3g, h). These data support the model that Tip60 and the β -catenin–reptin complex can indeed be directly recruited to the *KAI1* promoter by p50. To prove further that these complexes require p50 on the *KAI1* promoter, we used NF- κ B p50-knockout cells and wild-type control cells. In the absence of the p50 DNA-binding factor, in p50-deleted cells, Tip60, β -catenin and reptin were not detected on the promoter (Fig. 3i). Indeed, coimmunoprecipitation experiments confirmed a specific interaction of each

protein *in vivo* (Fig. 3j). Next we measured the NF- κ B-driven transcriptional activity in response to IL-1 β with or without over-expressed β -catenin (Fig. 3k–m). The reduced NF- κ B activity is correlated with the increased expression of β -catenin (Fig. 3k, l); however, this suppression is promoter-specific, because *ICAM-1* was not downregulated by β -catenin expression (Fig. 3m). Thus, although there seems to be a correlation between a requirement for Tip60 and antagonistic regulation by β -catenin of a subset of NF- κ B target genes, we cannot exclude the possibility that the β -catenin effect might not be functionally relevant to strongly induced promoters such as *ICAM-1*.

LNCaP metastatic cells showed greater nuclear β -catenin staining than that of non-metastatic cells (Supplementary Fig. S3c). An increase in β -catenin expression in the nucleus did not change the localization of Tip60, indicating that antagonistic roles of β -catenin and Tip60 coactivator might be related to relative levels rather than to altered compartmentalization (Supplementary Fig. S3d, e).

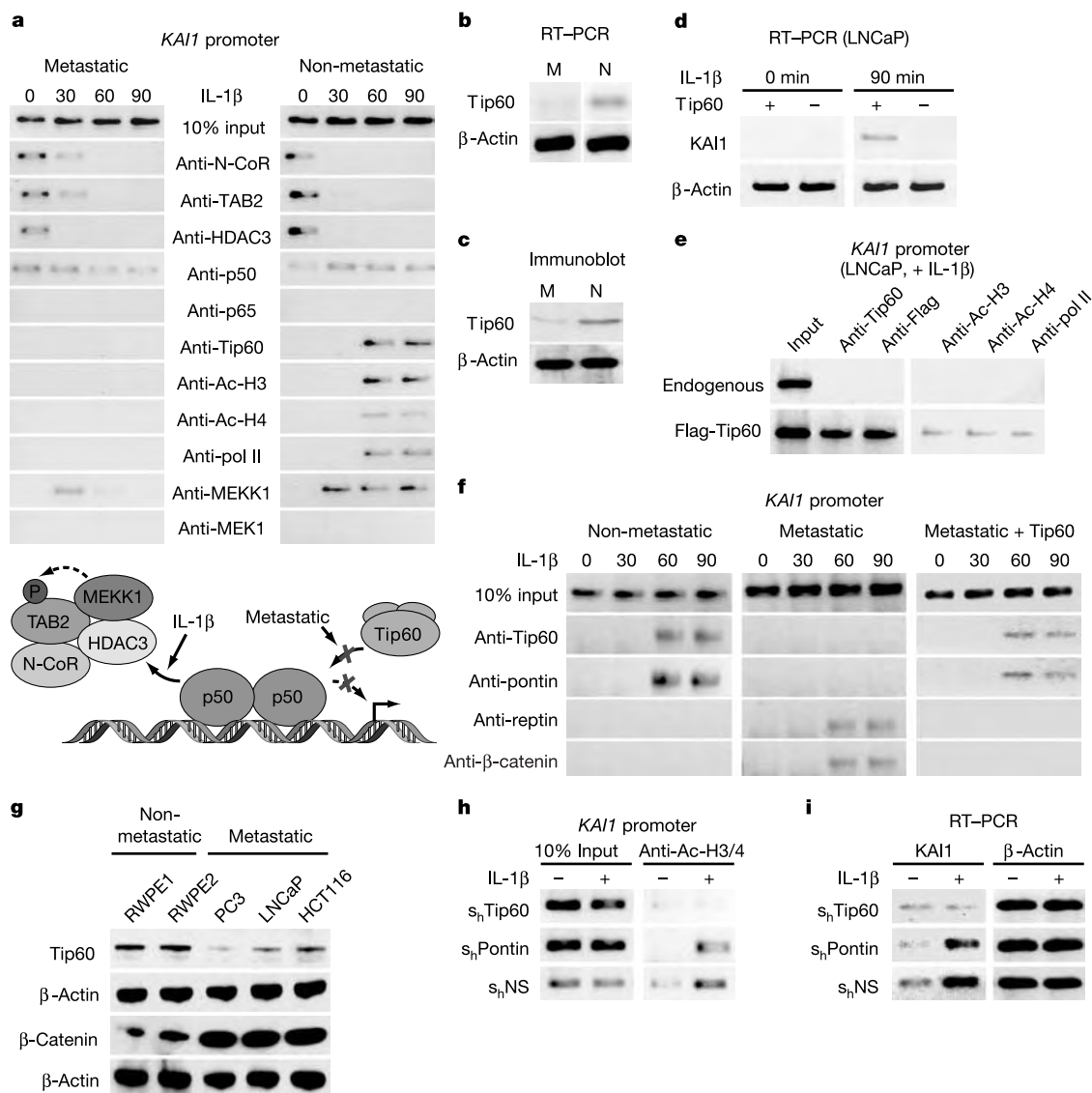


Figure 2 Failure of Tip60 recruitment on *KAI1* promoter is crucial for downregulation of *KAI1*. **a**, ChIP assay on *KAI1* promoter to compare cofactor dynamics in LNCaP and RWPE1 cells. Numbers above each gel are times in minutes. **b**, **c**, Analysis of Tip60 transcript (**b**) and protein expression (**c**). M, metastatic cells; N, non-metastatic cells. **d**, Overexpression of Tip60-induced *KAI1* expression in IL-1 β -treated LNCaP cells. **e**, **f**, ChIP analysis of Tip60-associated factors on *KAI1* promoter in RWPE1, LNCaP and

LNCaP cells stably overexpressing exogenous Tip60. Numbers above each gel in **f** are times in minutes. **g**, Immunoblot analysis of Tip60 and β -catenin in various cell lines. **h**, **i**, The sRNA-coupled ChIP assay (**h**) and *KAI1* mRNA induction (**i**) in the presence of sRNA against Tip60 (s_hTip60), pontin (s_hPontin) or non-specific sRNA-transfected RWPE1 cells (s_hNS).

To examine further whether β -catenin–reptin is directly required for the repressive function, we silenced its expression in LNCaP cells by using s_h RNA (Supplementary Fig. S4a, b). The s_h RNA coupled-ChIP assay in LNCaP cells revealed that knockdown of β -catenin alone resulted in failure of recruitment of either β -catenin or reptin on the promoter (Fig. 4a), which supports the model that binding to p50 seems to be mediated through β -catenin and not by interactions between p50 and reptin. Knockdown of reptin alone allowed the recruitment of β -catenin on the promoter, but with very weak acetylation signals (Fig. 4a). These data provide evidence that reptin is an important component of β -catenin-induced repression. Immunoblot analysis showed that knockdown of β -catenin

increased Tip60 expression to some extent (Fig. 4b). Knockdown of β -catenin or reptin restored the transcriptional activation of *KAI1* in LNCaP cells in the presence of IL-1 β (Fig. 4c).

Next we tested whether the repressive function of β -catenin is actually conferred by reptin, because knockdown of reptin alone allowed weak histone acetylation signals (Fig. 4a). We treated cells with trichostatin A (TSA), an established inhibitor of histone deacetylase (HDAC), and examined whether HDAC components are involved in reptin-mediated repression. Treatment of cells with TSA largely reduced the repressor function of Gal4–reptin (Fig. 4d). Single-cell microinjection of specific IgG of anti-HDAC1 relieved the repression by Gal4–reptin, indicating the likely functional

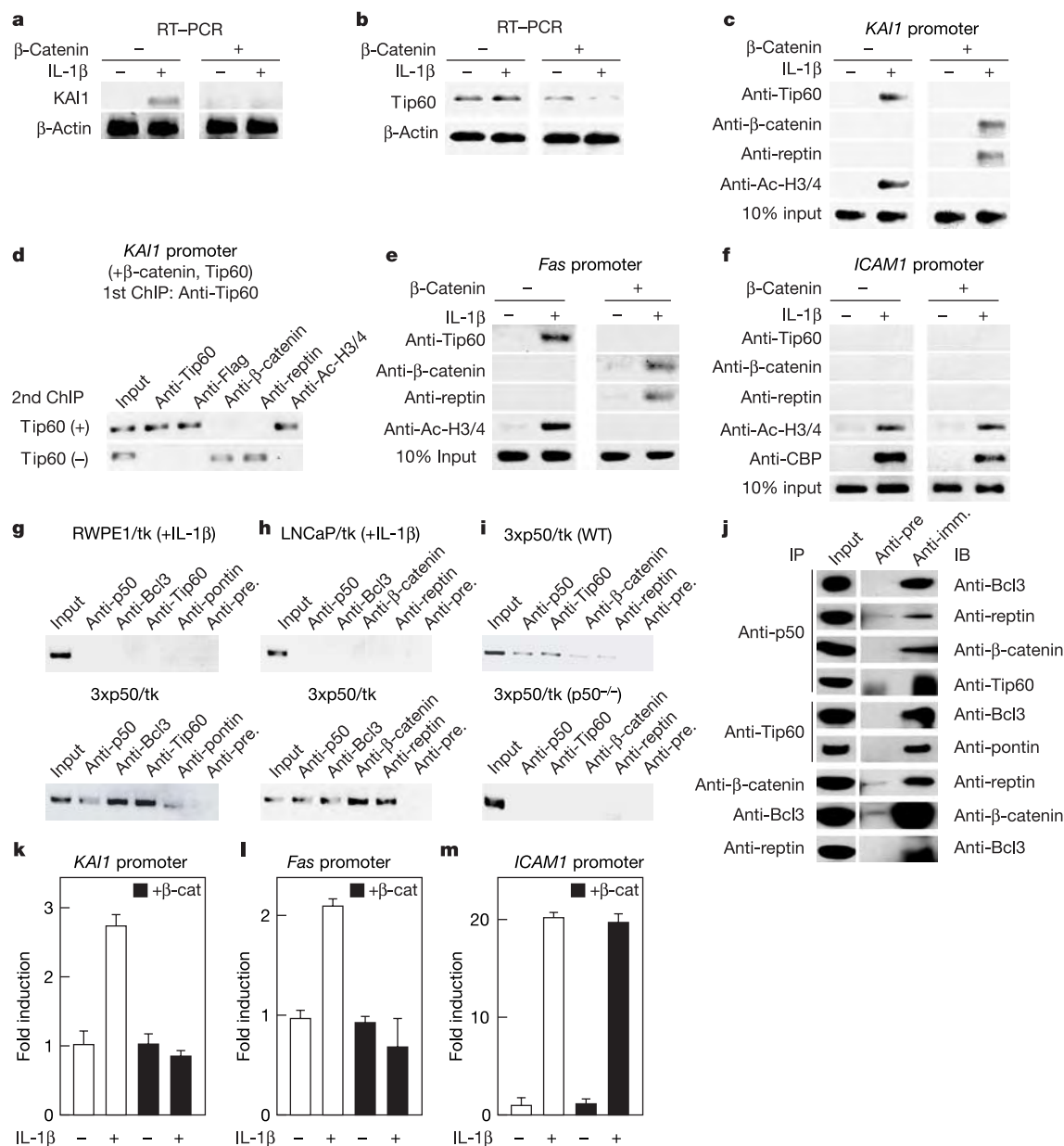


Figure 3 β -Catenin downregulates a subset of NF- κ B target genes along with reptin. **a, b**, Overexpression of β -catenin diminishes the *KAI1* (**a**) and *Tip60* (**b**) transcript in IL-1 β -treated RWPE1 cells. **c**, ChIP assay on the *KAI1* promoter in RWPE1 cells stably overexpressing β -catenin. **d**, Two-step ChIP assay in RWPE1 cells stably expressing β -catenin after transient transfection of Flag–Tip60. **e, f**, *Fas* promoter (**e**) was repressed by the overexpression of β -catenin in IL-1 β -treated RWPE1 cells, whereas *ICAM-1* promoter (**f**) was not. CREB binding protein (CBP) histone acetyltransferase was recruited

on *ICAM-1* promoter. **g–i**, Ability of p50 to recruit co-regulator complexes by promoter ChIP assay in RWPE1 (**g**), LNCaP (**h**) and p50 wild-type (WT) and knockout (**i**) cells. **j**, Coimmunoprecipitation assays to verify cofactor interactions. IP, immunoprecipitation; IB, immunoblot. **k–m**, Measurement of *KAI1* (**k**), *Fas* (**l**) and *ICAM-1* (**m**) promoter activity in the presence or absence of the overexpressed β -catenin. Error bars indicate standard deviations.

requirement of HDAC1 for the repression in those cells (Fig. 4e). Using proven *s*_hRNAs, we found that the ablation of HDAC1 with *s*_hRNA relieved the repression function of Gal4-reptin, whereas *s*_hRNA against HDAC3 had no effect (Supplementary Fig. S4c–e),

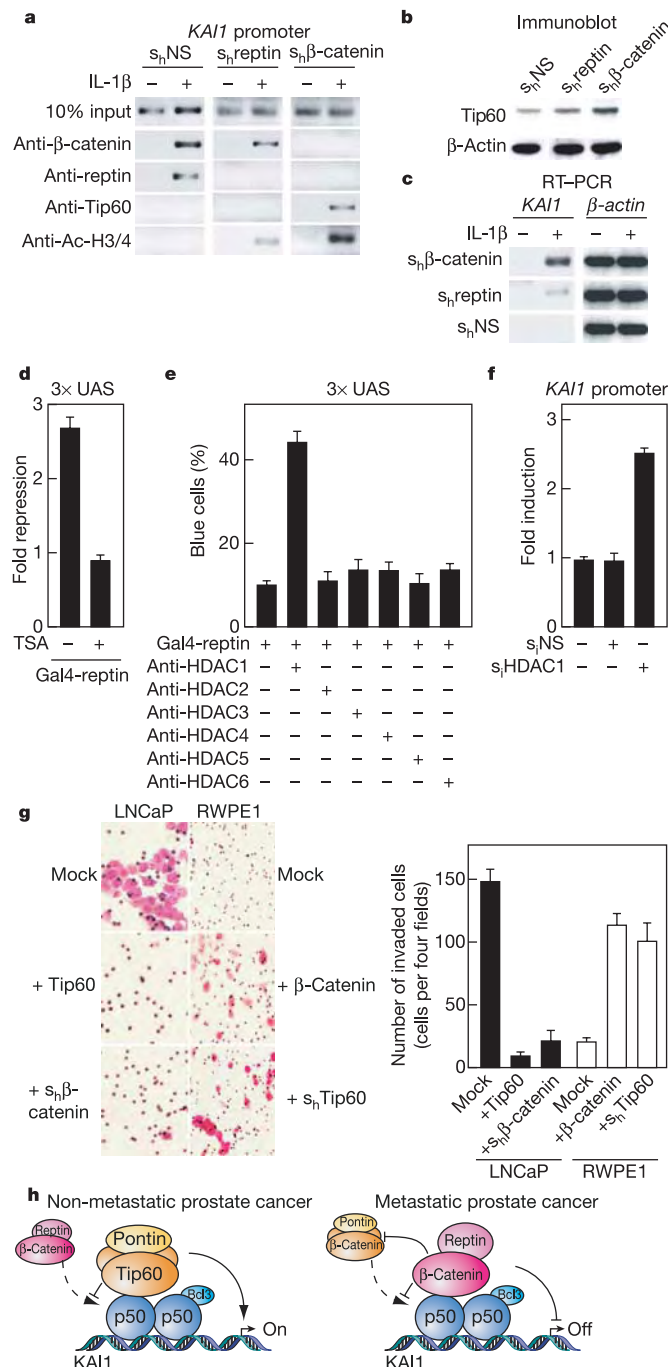


Figure 4 Tip60 and β-catenin switch affects the metastatic potential. **a–c**, The *s*_hRNA-coupled ChIP assay (**a**), immunoblot analysis of Tip60 (**b**) and *KAI1* message level (**c**) after knock-down of either β-catenin, reptin or non-specific *s*_hRNA in LNCaP cells. **d**, Luciferase reporter assay of Gal4-reptin on the Gal4-driven UAS/tk luciferase construct. UAS, upstream activating sequence element. **e**, Microinjection of IgGs against HDAC1 largely relieves the repression by Gal4-reptin. **f**, Co-transfection of *s*_hRNA against HDAC1 relieves the repression of *KAI1* promoter reporter about 2.5-fold in LNCaP cells. **g**, The invasive activity of LNCaP and RWPE1 cells assayed in Matrigel chambers. **h**, Schematic model of alternative recruitment of Tip60 coactivator and β-catenin–reptin repressor as a key regulatory switch for metastatic potential. Error bars indicate standard deviations.

indicating selective HDAC requirements for the repression function of reptin. Pull-down experiments with glutathione S-transferase confirmed the binding of reptin with HDAC1, and immunoprecipitation assays *in vivo* revealed the region of reptin for binding to HDAC1 (Supplementary Fig. S4f–h). Further, we found that *s*_hRNA against HDAC1 activated *KAI1* promoter reporter about 2.5-fold in LNCaP cells (Fig. 4f). The repression function of β-catenin–reptin is therefore conferred, at least in part, by HDAC1 activity.

Finally, we tested whether altering the Tip60 and β-catenin ratio would actually modulate the metastatic potential of invasive prostate cancer cells by using the Matrigel invasion assay (Fig. 4g, and Supplementary Fig. S5)^{23,24}. We therefore measured the ability of cells to traverse a Matrigel-coated membrane with 8-μm pores, a correlate of metastatic potential *in vivo*. As shown in Fig. 4g, overexpression of either Tip60 or *s*_hRNA against β-catenin in IL-1β-treated LNCaP cells decreased Matrigel invasion in comparison with the IL-1β-treated control cells, whereas RWPE1 cells expressing β-catenin expression vector or vector-based *s*_hRNA against Tip60 increased the Matrigel invasion. These data indicate that changes in the level of Tip60 and β-catenin expression might indeed be able to affect the metastatic potential.

Together our data reveal an unexpected pathway of transcriptional regulation of the metastasis suppressor gene *KAI1*. As the levels of nuclear β-catenin expression increase in the cell, the ability to selectively recruit a β-catenin–reptin complex to a specific subset of promoters in association with histone deacetylases confers a repressive function on β-catenin. This repressive function of β-catenin–reptin occurs in a promoter context-dependent fashion, with a subset of NF-κB target genes, including *KAI1*, being down-regulated by β-catenin–reptin. Conversely, the transcriptional activity of *KAI1* requires a sufficiently high level of Tip60 coactivator expression, which, directly or indirectly, is negatively regulated by increased levels of β-catenin. Because an increase in Tip60 expression is sufficient to restore the transcriptional activity of the *KAI1* promoter in metastatic cancer cells, the balance between the antagonistic actions of β-catenin–reptin components that mediate the repressive status and the activating properties of the Tip60 coactivator complex must be responsible for the dynamic process of regulation of a subset of NF-κB target genes, including the tumour metastasis suppressor *KAI1* (Fig. 4h). These studies thus indicate that crosstalk between the Wnt/β-catenin and the NF-κB pathway might be crucial for tumorigenesis and metastatic behaviour in cancer²⁵. Our findings, namely that cell-type-specific and promoter-specific differences in co-regulator recruitment are crucial in determining the downregulation of a subset of NF-κB genes, offer a paradigm for investigating mechanisms of downregulation of metastasis suppressor genes in the progression of metastatic behaviour in specific cancers. □

Methods

Materials and reagents

The following antibodies were obtained from Santa Cruz Biotechnology and previously described¹²: anti-HDAC1, anti-HDAC2, anti-HDAC3, anti-HDAC4, anti-HDAC5, anti-HDAC6, anti-p50, anti-p65, anti-MEKK1, anti-MEK1, anti-Bcl3, anti-β-catenin and anti-TAB2. Anti-TAB2 IgG is also from Affinity BioReagents. The following commercially available antibodies were used: anti-Tip60, anti-acetylated histone H3 and anti-acetylated histone H4 (Upstate Biotechnology) and anti-RNA polymerase II antibodies (Berkeley Antibody Company). Human IL-1β was purchased from Calbiochem, and TSA was from Sigma. Anti-reptin and anti-pontin IgGs were from O. Huber.

In vivo metastasis assay

LNCaP cells stably expressing either empty vector (pLNCX2; BD Biosciences) or *KAI1* expression vector (*KAI1*-pLNCX2) with luciferase reporter were injected into mouse prostate (10⁶ cells per mouse). The level of prostate-specific antigen (PSA) in these xenograft mice was monitored every week and luciferin (5 mg kg⁻¹ mouse body weight) was injected intravenously into a mouse when serum PSA level reached 15 ng ml⁻¹. Each mouse received 100 μl of luciferin solution intraperitoneally; 5 min after injection, the images were taken. The lung was dissected from each mouse and luciferase activity was recorded. Each tumour in prostate was also weighed as control.

ChIP, two-step ChIP and s_{p} RNA-coupled ChIP assay

The ChIP assay was conducted as described previously¹², with an average size of sheared fragments of about 300 base pairs to 1 kilobase. For PCR, 1 μ l from 50 μ l DNA extraction and 25–30 cycles of amplification were used. The following primers were used: *ICAM-1* promoter sense strand 5'-ACCTTAGCGCGGTGTAGACC-3', antisense strand 5'-CTCCGGACAAATGCTGC-3'; *Fas* promoter sense strand 5'-TCGAGGTCTCACC TGAAGT-3', antisense strand 5'-GAGACGAGCTCACGAAAAGC-3'; *KAI1* promoter sense strand 5'-ACCGTTAGGACGCGCGTGTAG-3', antisense strand 5'-CTTGGGAA GGCGGTGCGCTC-3'; 3 $\times$ p50/*tk* luciferase sense strand 5'-GACGCCAAAAACATAAA GAAAGGCC-3', antisense strand 5'-TTCACGTTCATTATAAATGTCGTTTC-3'. For the two-step ChIP assay, components were eluted from the first immunoprecipitation reaction by incubation with 10 mM dithiothreitol at 37 °C for 30 min and diluted 1:50 in ChIP dilution buffer (20 mM Tris-HCl pH 8.1, 150 mM NaCl, 2 mM EDTA, 1% Triton X-100) followed by reimmunoprecipitation with the second antibodies. Two-step ChIP was performed in essentially the same way as the first immunoprecipitations. For the s_{p} RNA-coupled ChIP assay, LNCaP or RWPE1 cells stably expressing each s_{p} RNA vector were selected, and cells stably expressing each s_{p} RNA vector were transiently transfected with each s_{p} RNA vector again to increase the knockdown effect. The cells were harvested and the immunoprecipitated chromatin was analysed by PCR with primers specific to the promoters.

Reporter assays

Luciferase activity was measured in a luminometer 48 h after transfection and normalized to β -galactosidase expression with a luciferase assay system (Promega). A single-cell microinjection assay was performed as previously described¹². Each experiment was performed on three independent coverslips consisting of 1,000 cells, with more than 300 cells injected. When no experimental antibody was used, preimmune rabbit or guinea pig IgG was injected at the same time, allowing the unambiguous identification of injected cells as well as serving as a preimmune control. Values are expressed as means $\pm$ standard deviations for at least three independent experiments.

RNA interference by s_{p} RNA

Target sequences for small hairpin RNAs for β -catenin, repton, pontin and Tip60 are as follows: β -catenin s_{p} RNA, 5'-GUCCUGUAUGAGUGGGAAC-3'; repton s_{p} RNA1, 5'-GGAGGAGACGGAGAUCAUC-3'; repton s_{p} RNA4, 5'-GAAGAUGUGGAGAUAGUG-3'; pontin s_{p} RNA1, 5'-GUUACUCAACUGAGAUA-3'; pontin s_{p} RNA2, 5'-GGUGAA GUCACAGAGCUAA-3'; Tip60 s_{p} RNA2, 5'-GGGACCAUCUCCUUCUUU-3'. Small interfering RNA pool for HDAC1, HDAC3, and non-specific control (SMARTpool HDAC1 and HDAC3) were purchased from Dharmacon.

Matrigel invasion assay

LNCaP cells stably overexpressing Tip60 or vector-based s_{p} RNA against β -catenin and RWPE1 cells stably overexpressing β -catenin or vector-based s_{p} RNA against Tip60 were used in the Matrigel invasion assay along with each control cell. Cultured cell line was pretreated with 5 ng ml⁻¹ IL-1 β for 24 h, and 2.5×10^4 LNCaP cells or 1.25×10^4 RWPE1 cells were loaded into the top of a 24-well Matrigel invasion chamber assay plate (BD Biocoat; BD Biosciences). Conditioned RPMI1640 medium containing 15% fetal bovine serum was added to the bottom chamber as a chemoattractant. After incubation for 22 h, the cells that had migrated to the lower chamber of the filter were fixed with 100% methanol, stained with Giemsa and quantified by counting the total number of cells in four independent areas. PC-3 cells were transfected with Tip60 expression plasmid and EGFP (enhanced green fluorescent protein) plasmid. After transfection, half of the cells were pretreated with IL-1 β for 24 h. Transfected cells (1.5×10^4) with or without IL-1 β treatment were loaded into the upper compartment of the Matrigel chamber; 10% fetal calf serum was put into the lower compartment of the chamber as a chemoattractant. All experimental studies were performed in accordance with the manufacturer's protocols. Similar data were obtained from three repeated experiments, and values are expressed as means $\pm$ standard deviations for at least three independent experiments.

Received 13 December 2004; accepted 3 February 2005; doi:10.1038/nature03452.

- Shevde, L. A. & Welch, D. R. Metastasis suppressor pathways—an evolving paradigm. *Cancer Lett.* **198**, 1–20 (2003).
- Steeg, P. S. Metastasis suppressors alter the signal transduction of cancer cell. *Nature Rev. Cancer* **3**, 55–63 (2002).
- Feng, Y., Lee, N. & Fearon, E. R. Tip49 regulates β -catenin-mediated neoplastic transformation and T-cell factor target gene induction via effects on chromatin remodeling. *Cancer Res.* **63**, 8726–8734 (2003).
- Bauer, A., Huber, O. & Kemler, R. Pontin52, an interaction partner of β -catenin binds to the TATA box binding protein. *Proc. Natl Acad. Sci. USA* **95**, 14787–14792 (1998).
- Bauer, A. *et al.* Pontin52 and Reptin52 function as antagonistic regulators of β -catenin signaling activity. *EMBO J.* **19**, 6121–6130 (2000).
- Wood, M. A., McMahon, S. B. & Cole, M. D. An ATPase/helicase complex is an essential cofactor for oncogenic transformation by c-Myc. *Mol. Cell* **5**, 321–330 (2000).
- Rottbauer, W. *et al.* Reptin and pontin antagonistically regulate heart growth in zebrafish embryos. *Cell* **111**, 661–672 (2002).
- Ikura, T. *et al.* Involvement of the Tip60 histone acetylase complex in DNA repair and apoptosis. *Cell* **102**, 463–473 (2000).
- Petrylak, D. P. Metastases suppressors and prostate cancer. *Nature Med.* **1**, 739–740 (1995).
- Dong, J.-T. *et al.* KAI1, a metastasis suppressor gene for prostate cancer on human chromosome 11p11.2. *Science* **268**, 884–886 (1995).
- Dong, J.-T. *et al.* Down-regulation of the KAI1 metastasis suppressor gene during the progression of human prostatic cancer infrequently involves gene mutation or allelic loss. *Cancer Res.* **56**, 4387–4390 (1996).

- Baek, S. H. *et al.* Exchange of N-CoR corepressor and Tip60 coactivator complexes links gene expression by NF- κ B and β -amyloid precursor protein. *Cell* **110**, 55–67 (2002).
- Elbashir, S. M. *et al.* Duplexes of 21-nucleotide RNAs mediate RNA interference in cultured mammalian cells. *Nature* **411**, 494–498 (2001).
- Hannon, G. J. RNA interference. *Nature* **418**, 244–251 (2002).
- Aberle, H., Bauer, A., Stappert, J., Kispert, A. & Kemler, R. β -Catenin is a target for the ubiquitin-proteasome pathway. *EMBO J.* **16**, 3797–3804 (1997).
- Orford, K., Crockett, C., Jensen, J. P., Weissman, A. M. & Byers, S. W. Serine phosphorylation-regulated ubiquitination and degradation of β -catenin. *J. Biol. Chem.* **272**, 24735–24738 (1997).
- Cheshire, D. R. & Isaacs, W. B. β -Catenin signaling in prostate cancer: an early perspective. *Endocr. Relat. Cancer* **10**, 537–560 (2003).
- Clever, H. Wnt breakers in colon cancer. *Cancer Cell* **5**, 5–6 (2004).
- Moon, R. T., Bowerman, B., Boutros, M. & Perrimon, N. The promise and perils of Wnt signaling through β -catenin. *Science* **296**, 1644–1646 (2002).
- Chesire, D. R., Ewing, C. M., Gage, W. R. & Isaacs, W. B. *In vitro* evidence for complex modes of nuclear β -catenin signaling during prostate growth and tumorigenesis. *Oncogene* **21**, 2679–2694 (2002).
- Deng, J. *et al.* β -Catenin interacts with and inhibits NF- κ B in human colon and breast cancer. *Cancer Cell* **2**, 323–334 (2002).
- Li, J. *et al.* Novel NEMO/I κ B kinase and NF- κ B target genes at the pre-B to immature B cell transition. *J. Biol. Chem.* **276**, 18579–18590 (2001).
- Albini, A. *et al.* A rapid *in vitro* assay for quantiating the invasive potential of tumor cells. *Cancer Res.* **47**, 3239–3245 (1987).
- Kobayashi, H. *et al.* Inhibition of *in vitro* ovarian cancer cell invasion by modulation of urokinase-type plasminogen activator and cathepsin B. *Cancer Res.* **52**, 3610–3614 (1992).
- Madrid, L. V. & Baldwin, A. S. Jr Regulation of NF- κ B by oncoproteins and tumor suppressor protein. *Methods Mol. Biol.* **223**, 523–532 (2003).

Supplementary Information accompanies the paper on www.nature.com/nature.

Acknowledgements We thank A. Hoffman for providing p50-knockout immortalized 3T3 cells; E. H. Koo and colleague for making KAI1-expressing LNCaP cells; D. A. Galloway for HPV-E6/E7 retroviral constructs; J.-T. Dong for KAI1 promoter clone; L. B. Owen-Schaub for *Fas* promoter reporter constructs; M.-C. Hung for mutant β -catenin constructs; K. Matsumoto for 293IL-R1 cells; K. I. Kim, Y. K. Park and J. M. Lee for critical reading; and J. Hightower and M. Fisher for figure and manuscript preparation. We acknowledge support from the National R&D programme for cancer control from Ministry of Health & Welfare, Korea Research Foundation grant, and the BK21 Research Fellowship from the Ministry of Education and Human Resources Development (S.H.B.), the NIH (M.G.R., C.L.S., D.W.R.), and the Prostate Cancer Foundation (PCF) and the National Cancer Institute (NCI) (M.G.R.). M.G.R. is an HHMI Investigator.

Competing interests statement The authors declare that they have no competing financial interests.

Correspondence and requests for materials should be addressed to S.H.B. (sbaek@snu.ac.kr) or M.G.R. (mrosenfeld@ucsd.edu).

Structure of the apoptotic protease-activating factor 1 bound to ADP

Stefan J. Riedl¹, Wenyu Li¹, Yang Chao¹, Robert Schwarzenbacher² & Yigong Shi¹

¹Department of Molecular Biology, Princeton University, Lewis Thomas Laboratory, Washington Road, Princeton, New Jersey 08544, USA

²Joint Center for Structural Genomics, University of California-San Diego, 9500 Gilman Drive, La Jolla, California 92093, USA

Apoptosis is executed by caspases, which undergo proteolytic activation in response to cell death stimuli¹. The apoptotic protease-activating factor 1 (Apaf-1) controls caspase activation downstream of mitochondria². During apoptosis, Apaf-1 binds to cytochrome *c* and in the presence of ATP/dATP forms an apoptosome, leading to the recruitment and activation of the initiator caspase, caspase-9 (ref. 2). The mechanisms underlying Apaf-1 function are largely unknown. Here we report the 2.2-Å crystal structure of an ADP-bound, WD40-deleted Apaf-1, which reveals the molecular mechanism by which Apaf-1 exists in an inactive state before ATP binding. The amino-terminal caspase

recruitment domain packs against a three-layered α/β fold, a short helical motif and a winged-helix domain, resulting in the burial of the caspase-9-binding interface. The deeply buried ADP molecule serves as an organizing centre to strengthen interactions between these four adjoining domains, thus locking Apaf-1 in an inactive conformation. Apaf-1 binds to and hydrolyses ATP/dATP and their analogues. The binding and hydrolysis of nucleotides seem to drive conformational changes that are essential for the formation of the apoptosome and the activation of caspase-9.

Apaf-1 exists in an inactive conformation in cells and requires ATP or dATP for activation³. The mechanistic role of ATP/dATP binding is unknown, although it was found to be essential for the formation of the apoptosome^{3–7}. The apoptosome, in turn, recruits and activates procaspase-9 and forms a holoenzyme with caspase-9 (ref. 8). Apaf-1 consists of an N-terminal caspase recruitment domain (CARD), a central nucleotide-binding oligomerization domain (NOD), and multiple WD40 repeats at the carboxy-terminal half. Apaf-1 is a representative member of the NOD family of proteins⁹. The WD40 repeats are thought to be responsible for binding to cytochrome *c* and have a regulatory role in Apaf-1 function^{10,11}.

To gain insights into Apaf-1 function, we crystallized the WD40-deleted Apaf-1 (residues 1–591) and determined its structure by multi-wavelength anomalous dispersion at 2.2 Å resolution (Fig. 1, Supplementary Table 1). The final atomic model of WD40-deleted Apaf-1 (hereafter referred to as Apaf-1) contains 586 amino-acid residues. The structure of Apaf-1 comprises five distinct domains—CARD, three-layered α/β domain, helical domain I, winged-helix domain and helical domain II—that stack against each other through extensive inter-domain interactions (Fig. 1a, b, Supplementary Fig. 1). Packing between these five domains gives rise to a relatively compact structure, with a dimension of 80 Å × 65 Å × 55 Å. The N-terminal CARD (residues 1–107) contains six α -helices, α 1– α 6, arranged in a Greek-key topology. The three-layered α/β domain (residues 108–284) consists of five parallel β -strands, β 1– β 5, in the centre, sandwiched by four α -helices on each side (Fig. 1). A short α -helical domain (helical domain I, residues 285–365), containing helices α 16– α 19, is followed by an unexpected winged-helix domain (residues 366–450), which usually appears in eukaryotic transcription factors¹². The extended helical domain (helical domain II, residues 451–586) is exclusively composed of α -helices, α 25– α 32, arranged in a right-handed super-helical conformation.

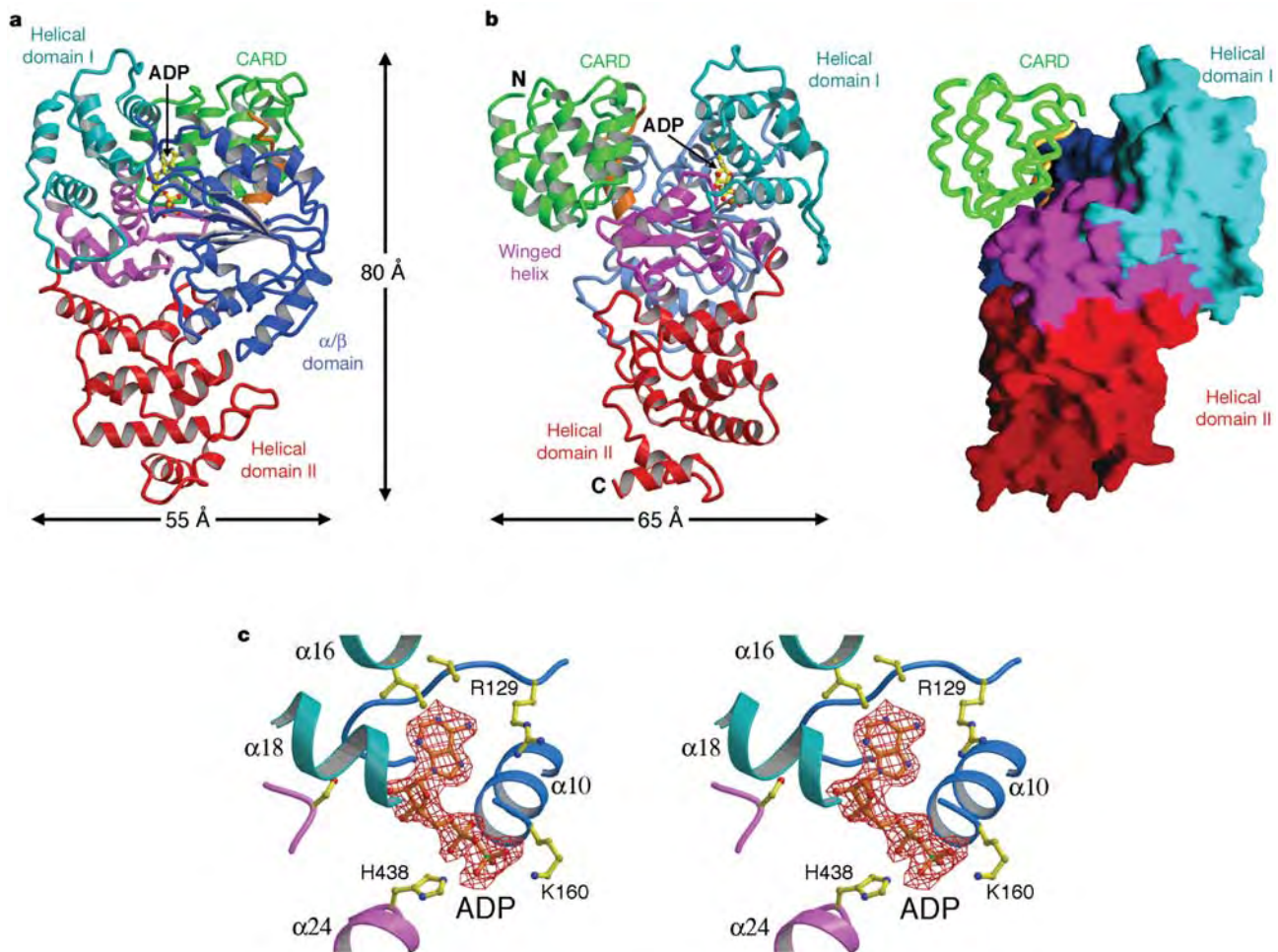


Figure 1 Overall structure of the WD40-deleted Apaf-1 bound to ADP. **a**, A ribbon diagram of the structure of Apaf-1 (residues 1–591) bound to ADP. Apaf-1 sequentially comprises five distinct domains: CARD (coloured green), an α/β fold (blue), helical domain I (cyan), a winged-helix domain (magenta) and helical domain II (red). These five domains pack against one another to generate a relatively compact structure. ADP binds to the hinge region between the α/β fold and helical domain I but is also coordinated by

two critical residues from the winged-helix domain. **b**, The structure of Apaf-1 in another orientation. Relative to **a**, Apaf-1 is rotated 90° along a vertical axis within the plane of paper. **c**, A stereo view showing the binding of ADP in the Apaf-1 structure. The $|F_o - F_c|$ omit electron density map, contoured at 2.0 σ , is calculated with the omission of ADP and shown in red around ADP. Figures 1–3 and 5 were prepared using MOLSCRIPT²⁹ and GRASP³⁰.

Although neither ATP nor any other nucleotide was supplemented during purification or crystallization, the electron density map unambiguously identified ADP as the bound nucleotide (Fig. 1c). ADP is bound at the interface between three domains: the α/β fold, helical domain I and the winged-helix domain (Fig. 1c).

We searched for structural homologues of Apaf-1 with the DALI server¹³. Two of the most similar structures were found to be the D2 hexamerization domain of *N*-ethylmaleimide-sensitive fusion protein (NSF), which is an essential ATPase required for intracellular vesicle fusion¹⁴, and p97 (ref. 15). Both NSF and p97 belong to the AAA (for ATPases associated with various cellular activities) family of ATPases¹⁶. The conserved regions encompass the entire α/β fold and helical domain I. The presence of a short helical domain following the α/β fold is a hallmark of the AAA+ family of ATPases, in which the helical domain contributes to nucleotide binding energetically¹⁶. This analysis, in conjunction with sequence features (Supplementary Fig. 1) and an earlier report¹⁷, identifies Apaf-1 as a member of the AAA+ family of ATPases.

The CARD domain of Apaf-1 interacts with the prodomain of caspase-9, and this interaction is essential for the recruitment and subsequent activation of caspase-9 (refs 3, 18). In Apaf-1, the N-terminal CARD stacks against the α/β fold and the winged-helix domain through a large interface involving helices α 2, α 4 and α 5 of

the CARD domain (Fig. 2). Interestingly, helix α 2 of Apaf-1 CARD is also required for binding to the prodomain of caspase-9 (ref. 18). Docking of caspase-9 prodomain to the CARD domain of Apaf-1 resulted in significant steric clashes between the caspase-9 prodomain and the α/β fold of Apaf-1 (Fig. 2a). Thus, for Apaf-1 to interact with caspase-9, the relative orientation of CARD with respect to the rest of Apaf-1 structure must be changed. In addition, the packing of CARD with other domains of Apaf-1 further restricts access to the bound ADP molecule (Fig. 2b, c). These observations suggest that ADP-bound Apaf-1 adopts a closed conformation and that exchange and/or hydrolysis of the nucleotide is likely to weaken the interaction of CARD with other domains of Apaf-1. In agreement with this analysis, caspase-9 binds efficiently to full-length Apaf-1 only in the presence of cytochrome *c* and dATP/ATP³.

The interactions between CARD and the α/β fold and the winged-helix domain involve a network of 13 inter-domain hydrogen bonds and some van der Waals contacts, resulting in the burial of 2,550 Å² surface area (Fig. 2c). At one end of the interface, the bridging helix α 8 mediates four hydrogen bonds and several van der Waals contacts. In the centre of this interface, Glu 78 of CARD makes a bifurcated hydrogen bond to Arg 428 of the winged-helix domain. This interaction is buttressed on each end by one inter-domain hydrogen bond and a few van der Waals contacts (Fig. 2c).

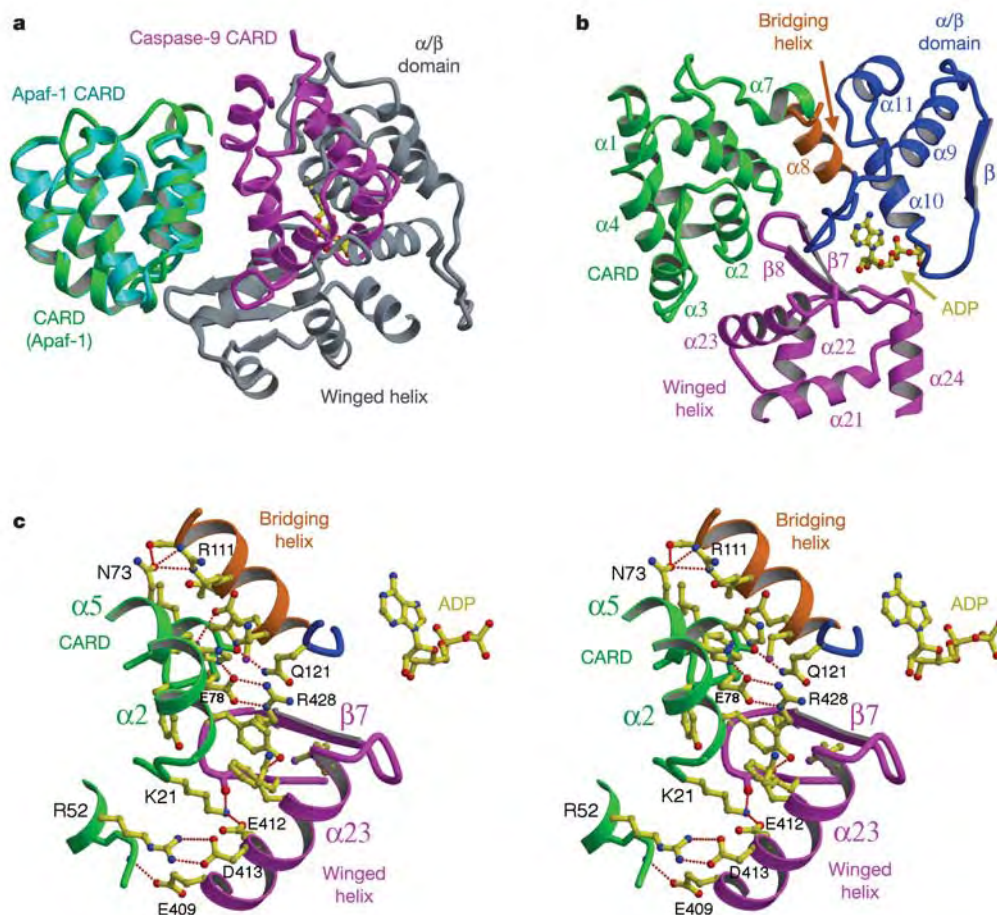


Figure 2 The CARD domain of Apaf-1 exists in a closed conformation. **a**, The structural elements of CARD required for binding to the prodomain of caspase-9 are involved in inter-domain interactions within Apaf-1. Docking the caspase-9 prodomain onto the CARD domain of the ADP-bound Apaf-1 causes severe steric clashes between the caspase-9 prodomain and the α/β fold and winged-helix domains of Apaf-1. The structure of Apaf-1 CARD (cyan) bound to the prodomain of caspase-9 (pink) was as described¹⁸. CARD of the ADP-bound Apaf-1 is coloured green, whereas the α/β fold and

winged-helix domain of Apaf-1 are shown in grey. **b**, A schematic diagram of the inter-domain packing between CARD (green), the α/β fold (blue) and the winged-helix domain (magenta). The bridging helix (gold), which forms a single folding unit with the α/β fold, closely stacks against helix α 5 of the CARD domain. **c**, A stereo view of the specific interactions between CARD and the α/β fold and winged-helix domains. Hydrogen bonds are represented by red dashed lines. ADP is shown to indicate the relative orientations of CARD and ADP. The colouring scheme in **b** and **c** is the same as in Fig. 1.

At the other end of the interface, Arg52 of CARD donates two specific hydrogen bonds to the carboxylate group of Asp 413 on the winged-helix domain, which is reinforced by three additional inter-domain hydrogen bonds. These interactions probably stabilize the contact between the α/β fold and the winged-helix domain.

To examine whether CARD in the closed conformation of Apaf-1 is capable of binding to caspase-9, we incubated Apaf-1 (residues 1–591) with the full-length processed caspase-9 and compared the behaviour of the complex with each component using gel filtration. The result indicates that Apaf-1 and caspase-9 form a 1:1 complex in the absence of ATP/dATP (Supplementary Fig. 2a). Next, we investigated whether this 1:1 complex had any impact on the catalytic activity of caspase-9. The activity of caspase-9 was significantly elevated about fourfold in the presence of Apaf-1 despite the absence of ATP/dATP (Supplementary Fig. 2b). This observation indicates that, even in a 1:1 complex, WD40-deleted Apaf-1 is able to enhance the catalytic activity of caspase-9 allosterically.

One surprising revelation of the Apaf-1 structure is that the bound nucleotide is ADP rather than ATP or dATP. The binding pocket for ADP is formed at the junction of four domains, CARD, α/β fold, helical domain I and winged-helix domain. Through extensive interactions, ADP seems to serve as an organizing centre to bring together these four adjoining domains and locks Apaf-1 in a closed conformation (Figs 1 and 3). Consequently, the bound ADP molecule is deeply buried and the only narrow channel from ADP to the solvent is blocked by the packing of the CARD domain (Fig. 3a, right). This structural organization indicates that the unpacking of CARD, which may be achieved through interaction

with the prodomain of caspase-9, might lead to a more accessible binding pocket for nucleotide. Indeed, the binding of ATP/dATP to Apaf-1 was shown to be facilitated by the presence of caspase-9 (ref. 5). Our structural analysis also suggests that perturbation to the nucleotide-binding pocket, such as exchange or hydrolysis of a nucleotide, might result in the disruption of the inter-domain packing between the four adjoining subunits.

The specific binding of ADP is achieved by a total of eight direct hydrogen bonds from residues in Apaf-1, including two to the adenine base and six to the phosphate groups (Fig. 3b). The N1 and N6 atoms of the adenine base are coordinated by the main-chain amide and carbonyl groups of Val 127, respectively. These two hydrogen bonds are specific for the adenine base but not the guanine, explaining why GTP/dGTP fails to activate Apaf-1 (refs 3, 5). The α -phosphate is coordinated by one hydrogen bond from the amide group of Val 162. The β -phosphate is coordinated by five hydrogen bonds, from the amide groups of Gly 159, Lys 160 and Ser 161, the side-chain amino group of Lys 160 and the imidazole group of His 438. Gly 159, Lys 160, Ser 161 and Val 162 come from the P-loop (also known as the Walker A motif) of the α/β fold, whereas His 438 is located within the winged-helix domain. Although ADP is bound in the structure, there seems to be sufficient room to accommodate the γ -phosphate of an ATP molecule.

A few well-ordered water molecules seem to be important in binding to ADP. The side chain of Arg 129 and the carbonyl group of Gly 159 make a water-mediated hydrogen bond to the N7 atom of the adenine base, whereas the carbonyl group of Val 125 and Ser 422 make water-mediated hydrogen bonds to the adenine base and the

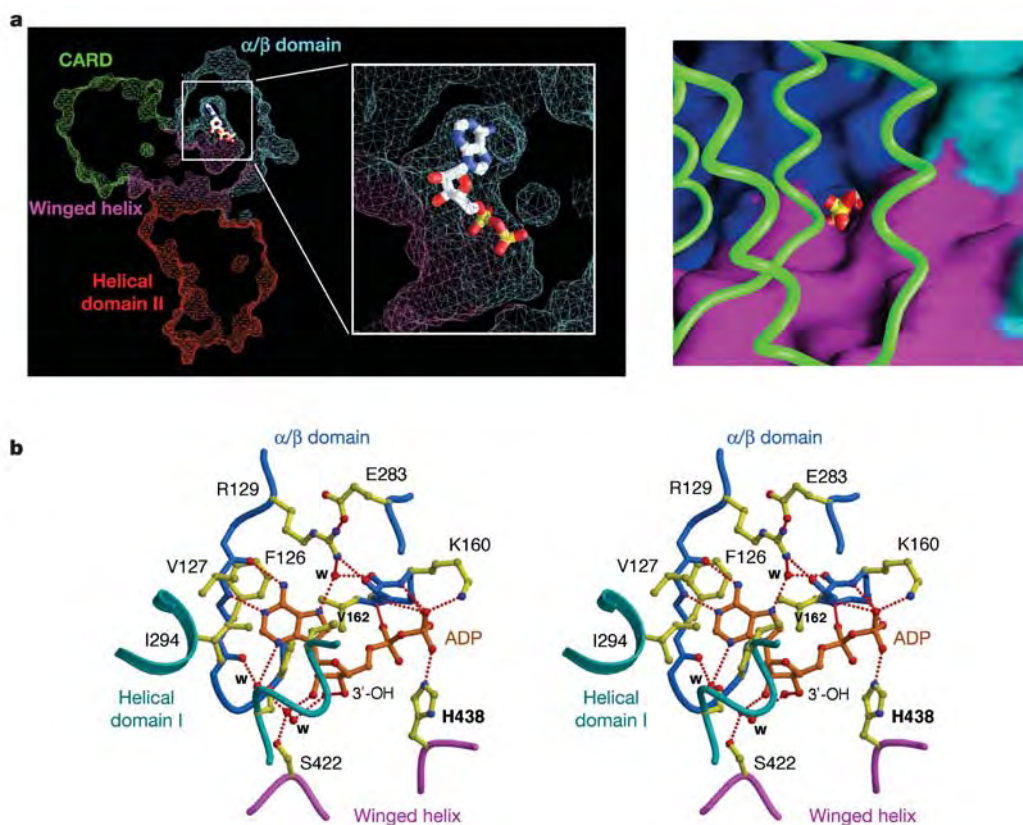


Figure 3 ADP serves as an organizing centre for the adjoining three domains and locks Apaf-1 in an inactive conformation. **a**, ADP is deeply buried and inaccessible to even small molecules unless the surrounding domains undergo conformational changes. Left, a cross-section of Apaf-1, with its van der Waals surface represented by a colour-coded mesh. Right, the CARD domain (green) blocks the only solvent channel to the ADP-binding

pocket. **b**, A stereo representation of the coordination of ADP by residues from three domains. As in other AAA+ ATPases¹⁶, ADP is bound primarily in the hinge region between the α/β fold (blue) and helical domain I (cyan). In Apaf-1, the winged-helix domain also contributes a direct hydrogen bond (from His 438) to the β -phosphate group and a water-mediated hydrogen bond (from Ser 422) to the ribose.

ribose, respectively (Fig. 3b). In addition to hydrogen bonds, several residues stabilize the adenine and the ribose moieties through van der Waals contact, including Pro 123, Phe 126, Val 127, Arg 129, Gly 159 and Val 162 from the α/β fold, and Ile 294, Pro 321, Leu 322 and Ser 325 from helical domain I.

In comparison with other AAA+ ATPases, the unique feature of Apaf-1 is the involvement of the winged-helix domain in the coordination of ADP, with His 438 and Ser 422 contributing two hydrogen bonds (Fig. 3b). Of particular interest is His 438, which replaces a characteristic feature of AAA+ ATPases named sensor II¹⁶, coordinates the β -phosphate group and thus directly links the winged-helix domain to the bound nucleotide. Consequently, in contrast with known structures of other AAA+ ATPases, the ADP molecule in Apaf-1 is deeply buried. The significant yet weak interactions between ADP and the winged-helix domain indicate that this domain might be prone to conformational changes.

The sequences of Apaf-1 contain all key elements of an ATPase that are required for activity, including the P-loop and the Walker B motif (Supplementary Fig. 1). In addition, although Apaf-1 has a higher affinity for ATP than for ADP^{5,19}, the bound nucleotide in

Apaf-1 is ADP, further indicating a possible ATPase activity. However, whether Apaf-1 is a genuine ATPase remains unresolved^{4,5,7}. To address this issue, we purified recombinant Apaf-1 protein (residues 1–591) to homogeneity. Using thin-layer chromatography (TLC), we reconstituted an *in vitro* ATPase assay to detect the hydrolysis of α -³²P-labelled ATP (Fig. 4a). Apaf-1, but not caspase-9, exhibited an ATPase activity, as shown by the appearance of ADP on TLC (Fig. 4a, lanes 1 and 2). Incubation of this reaction with EDTA resulted in the abrogation of the observed ATPase activity (Fig. 4a, lane 3), presumably because of the removal of the Mg²⁺ ion required for ATP hydrolysis. Supplementing Mg²⁺ in excess of EDTA restored the ATPase activity (Fig. 4a, lane 4). The presence of caspase-9 had no significant effect on the ATPase activity of Apaf-1 (Fig. 4a, lanes 5–7). These results show that Apaf-1 is capable of hydrolysing ATP and that this activity is dependent on Mg²⁺.

Next we examined whether the ATPase activity of Apaf-1 is essential to caspase-9 activation by using an assay reconstituted *in vitro* (Fig. 4b). As expected, the proteolytic activity of isolated caspase-9 was low (Fig. 4b, lane 1). Incubation of caspase-9 with

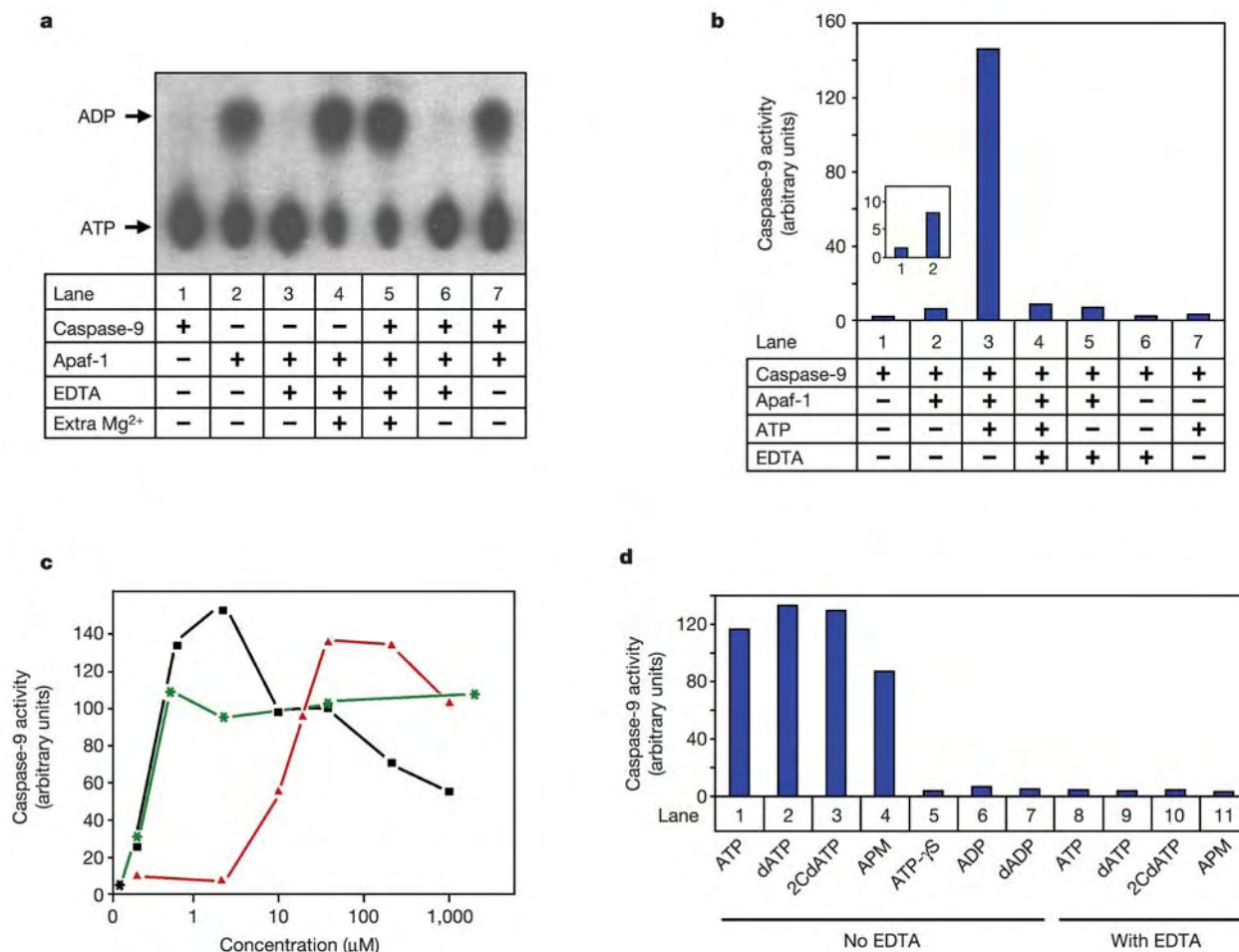


Figure 4 Apaf-1 is an active ATPase, and its ATPase activity seems to be essential to the activation of caspase-9. **a**, Apaf-1 hydrolyses ATP and its activity is blocked by the removal of magnesium. The ATPase activity was monitored by TLC. Apaf-1 (lane 2), but not caspase-9 (lane 1), hydrolysed [α -³²P]ATP into [α -³²P]ADP and phosphate. The chelation of the required Mg²⁺ ion by EDTA resulted in the abrogation of the ATPase activity (lane 3), whereas the addition of excess Mg²⁺ (over EDTA) restored the ATPase activity of Apaf-1 (lane 4). **b**, The ATPase activity of Apaf-1 is required for the activation of caspase-9. In the presence of 10 μ M ATP, Apaf-1 potentially activated caspase-9 (lane 3). Inhibition of the ATPase activity of Apaf-1 by EDTA treatment resulted in the abrogation of

ATP-dependent caspase-9 activation (lane 4). **c**, The optimal concentration ranges of nucleotides required for Apaf-1-mediated activation of caspase-9 are different. Unlike ATP (green), which exhibits a constant level of caspase-9 activation beyond 1 μ M, dATP (blue) has an optimal range of 1–5 μ M. Red, 2CdATP. **d**, Other hydrolysable ATP/dATP analogues can replace ATP to stimulate Apaf-1-mediated caspase-9 activation. Shown here are dATP (lane 2), 2CdATP (lane 3) and 2-thiomethyl-ATP (APM). Incubation with EDTA led to the abrogation of caspase-9 activation (lanes 9–11). The non-hydrolysable ATP analogue ATP- γ S does not have any effect on Apaf-1-mediated caspase-9 activation.

Apaf-1 significantly improved its catalytic activity about fourfold (Fig. 4b, lane 2); the addition of ATP further drastically enhanced caspase-9 activity by about two orders of magnitude (Fig. 4b, lane 3), confirming the critical function of ATP binding. Eliminating ATPase activity through the use of EDTA completely abolished the effect of ATP (Fig. 4b, lane 4), indicating that ATP hydrolysis might be essential to Apaf-1-mediated caspase-9 activation.

As previously reported^{3,19,20}, dATP and 2CdATP also promoted Apaf-1-mediated caspase-9 activation (Fig. 4c, d). As with ATP, the dATP or 2CdA-mediated effect seemed to depend on its hydrolysis, because the presence of 1 mM EDTA abolished caspase-9 activation (Fig. 4d, lanes 2, 3 and 9, 10). In addition, another ATP analogue, 2-methylthio-ATP, also supported Apaf-1-mediated caspase-9

activation, and incubation with EDTA abrogated this effect (Fig. 4d, lanes 4 and 11). The requirement of ATP/dATP hydrolysis for caspase-9 activation was further confirmed by the use of the non-hydrolysable analogues ATP- γ S (Fig. 4d, lane 5), AMP-PNP and AMP-PCP (data not shown), which were unable to support caspase-9 activation.

Among the nucleotide analogues that support Apaf-1-mediated caspase-9 activation, 2CdATP is particularly noteworthy because it is a metabolite of the anti-cancer drug cladribine (2CdA), whose cytotoxicity depends on the accumulation of 2CdATP. The observation that 2CdATP is capable of promoting Apaf-1-dependent caspase-9 activation suggests that other synthetic nucleotide analogues might be developed to enhance Apaf-1-mediated apoptosis

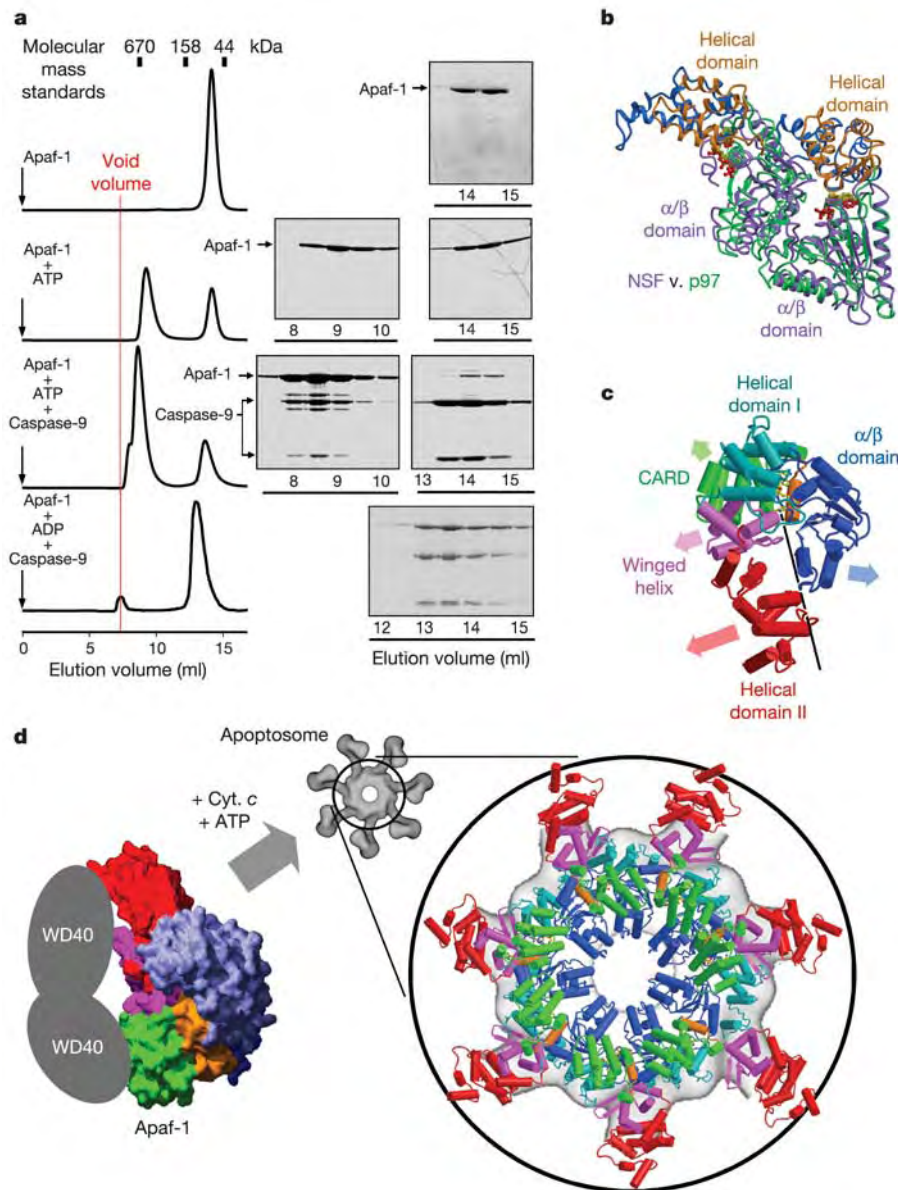


Figure 5 A model for the formation of the apoptosome. **a**, Apaf-1 forms an apoptosome in the presence of ATP. Apaf-1 (residues 1–591) was eluted as a monomer on gel filtration (top). In the presence of 10 μ M ATP, Apaf-1 forms a homo-oligomer (apoptosome, second from top). This apoptosome efficiently recruited caspase-9 to form a holo-enzyme (third from top). ADP has no effect on apoptosome formation (bottom). **b**, The interface for the oligomerization of AAA ATPases is conserved. For both NSF¹⁴ and p97 (ref. 15), the back corner of the α/β domain of one protomer packs against the front, triangular-shaped, surface of an adjacent protomer. The α/β and helical domains are coloured purple/orange

for NSF and green/blue for p97. **c**, Proposed conformational changes in Apaf-1. The conformational changes are induced by the binding and hydrolysis of a nucleotide. **d**, A model for the formation of the apoptosome. Left, a model of the full-length Apaf-1, in which the WD-40 repeats interact with the N-terminal domains. Centre and right, binding of cytochrome *c* and the binding and hydrolysis of ATP/dATP induce conformational changes that result in the formation of a heptameric apoptosome, which is mediated primarily by the α/β domain and helical domain I. This model is consistent with the 27-Å structure of the apoptosome²⁴.

in cancer cells. This possibility is supported by the enlarged ADP-binding pocket in Apaf-1 (Fig. 3a).

In response to a wide range of intrinsic cell death stimuli, Apaf-1 binds to the cytochrome *c* that is released from mitochondria, and in the presence of ATP or dATP Apaf-1 forms an oligomeric complex dubbed an apoptosome^{3,4,6,7,21}. Assembly of the apoptosome is required for full caspase-9 activation. We confirmed that Apaf-1 (residues 1–591) formed an apoptosome in the presence of ATP (Fig. 5a). Incubation of the apoptosome with caspase-9 resulted in the formation of a caspase-9 holoenzyme (Fig. 5a), which had a greatly enhanced catalytic activity compared with the isolated caspase-9 (Fig. 4b, lanes 1 and 3). In contrast, the caspase-9 holoenzyme is not formed in the presence of ADP (Fig. 5a) or in the absence of exogenous nucleotide (Supplementary Fig. 2a).

On the basis of structural analysis (Figs 1–3), we propose that the binding and hydrolysis of ATP at the junction of four domains result in the reorganization of these domains and the subsequent formation of the apoptosome. Indeed, the AAA+ family of proteins are known to couple ATP hydrolysis with conformational changes that have functional consequences¹⁶. To define ATP hydrolysis-induced conformational changes in Apaf-1, we studied known AAA+ ATPases and found that their oligomerization is mediated by a conserved mode of domain organization. For example, for both NSF (ref. 14) and p97 (ref. 15), the far end of the α/β fold of one protomer stacks against the wedge between the α/β fold and its C-terminal helical domain of the adjacent protomer (Fig. 5b). This interaction is repeated six times, resulting in the hexamerization of NSF. The same interface topology was also observed in a number of other AAA+ ATPases, including the bacterial protein HslU²² and the viral protein SV40 large T antigen²³. For Apaf-1 to adopt a similar arrangement the CARD domain has to be moved, and the winged-helix domain and helical domain II must be dislocated from their current positions and rotated by about 45–50° (Fig. 5c). This domain reorganization would allow the formation of a heptamer for Apaf-1 (Fig. 5d), which seems to resemble the core of the apoptosome as revealed by cryo-electron microscopy²⁴.

The proposed conformational changes in Apaf-1 are in agreement with the structural observation of the four adjoining domains at the ADP-binding site. In addition, biochemical characterization indicates that ATP hydrolysis is coupled to apoptosome formation and subsequent caspase-9 activation (Figs 4 and 5a). This model is speculative, and determination of the detailed structural organization of the apoptosome awaits a structure at an atomic resolution.

Thus, the structure of WD40-deleted Apaf-1 bound to ADP reveals the underlying molecular mechanism by which Apaf-1 maintains itself in an inactive state before ATP binding and hydrolysis. ADP is essential in that it serves as an organizing centre to bridge interactions between four adjoining domains, which locks Apaf-1 in an inactive state. It remains to be seen whether this mechanism is generally applicable to other members of the NOD family of proteins. This structure serves as a framework for understanding other functions of Apaf-1 as well as for understanding the functions of Apaf-1 homologues in other species. □

Methods

Protein preparation

All constructs for protein expression were generated with a standard PCR-based cloning strategy. Apaf-1 (residues 1–591, in pET29) was overexpressed in *Escherichia coli* strain BL21(DE3). The soluble fraction of Apaf-1 was purified from the cell lysate with a Ni²⁺-nitrilotriacetate (Qiagen) column and further fractionated by an anion-exchange column (Source-15Q; Pharmacia). Recombinant full-length caspase-9, also overexpressed in pET29, was purified in a similar manner with an additional gel-filtration chromatography step (Superdex-200; Pharmacia).

Crystallization

Purified protein from anion-exchange chromatography (3 mg ml⁻¹ Apaf-1 in 20 mM HEPES pH 7.8, 2 mM dithiothreitol (DTT), 250 mM NaCl) was used directly for crystallization. Crystals were produced by using the sitting-drop vapour-diffusion

method. Protein samples (2 μ l) were mixed with an equal volume of reservoir solution containing 100 mM HEPES pH 7.1, 250 mM ammonium acetate, 19% (w/v) PEG-3350 and 1 μ l of 100 mM DTT. Crystals appeared after 1–3 weeks at 4 °C, with a typical size of 0.1 $\times$ 0.1 $\times$ 0.1 mm³. Derivative crystals were obtained by soaking crystals overnight in mother liquor containing 1 mM mercury thimerosal followed by back-soaking for 5 min. Native and derivative crystals were then equilibrated in a cryo-protectant buffer containing well buffer plus 5% glycerol and were flash-frozen in liquid nitrogen. Derivative crystals belong to the space group *P*2₁, with unit cell dimensions *a* = 47.83 Å, *b* = 76.01 Å, *c* = 81.15 Å, β = 91.33°, and contain one protein molecule per asymmetric unit. Native crystals belong to the space group *P*1, with unit cell dimensions *a* = 75.96 Å, *b* = 92.88 Å, *c* = 94.99 Å, α = 62.96°, β = 89.99° and γ = 90.05°, and contain four molecules of Apaf-1 per asymmetric unit (the unit cell).

Data collection and structure determination

Anomalous diffraction data were collected at the NSLS beamline X25 by using three wavelengths corresponding to the inflection point, the high-energy remote and the peak of a Hg-MAD (multi-wavelength anomalous dispersion) experiment. In addition, a 2.2-Å native data set was collected at the CHESS beamline A1. Data were processed with HKL2000 (ref. 25). The initial structure was determined with 3.1-Å Hg-MAD data by using SOLVE/RESOLVE²⁶. Model building and TLS (translation-libration-screw) refinement were performed with the 2.2-Å native data set by using O²⁷ and REFMAC5 (ref. 28). Tight and medium NCS (non-crystallographic symmetry) restraints were imposed for main chain and side chain atoms, respectively. Data collection and refinement statistics are summarized in Supplementary Table 1. The final atomic model includes four monomers (chains A, B, C and D; residues 1–586), four ADP molecules and 782 water molecules in the asymmetric unit. No electron density was observed for residues 587–591 or for residues 95–103 in chains A and B.

ATPase assay

ATPase activity was determined with a TLC assay designed for the quantitative measurement of the conversion of [α -³²P]ATP to ADP. Reactions were initiated by mixing Apaf-1 (residues 1–591, final concentration 2 μ M) with a solution containing 20 mM HEPES pH 7.5, 10 mM DTT, 10 μ M MgCl₂, 0.1 μ M [α -³²P]ATP and 8 μ M ATP. The reactions were incubated at 22 °C and then quenched by the addition of equal volumes of developing solvent (1 M formic acid, 0.5 M LiCl). Samples were evaluated on the basis of the different mobilities of ATP and ADP on TLC with polyethyleneimine-cellulose F plates.

Caspase-9 activation assay

A fluorogenic assay was developed to monitor caspase-9 activation by Apaf-1 (residues 1–591). ATP or another nucleotide or its analogue was mixed with buffer (20 mM HEPES pH 7.5, 100 mM KCl, 5 mM DTT) to a final concentration of 1 mM in the presence or absence of 1 mM EDTA. Then, as indicated, caspase-9 and Apaf-1 were added to a final volume of 200 μ l and final concentrations of 0.2 and 2 μ M, respectively. After incubation for 5 min at room temperature, the caspase-9 substrate LEHD coupled to 7-amino-4-methylcoumarin was added to the reaction at a final concentration of 200 μ M. The conversion of the fluorogenic substrate was then measured in a Hitachi F2500 fluorescence spectrophotometer with an excitation wavelength of 380 nm and monitoring of the fluorescence at 440 nm.

Received 5 December 2004; accepted 16 February 2005; doi:10.1038/nature03465.

- Riedl, S. J. & Shi, Y. Molecular mechanisms of caspase regulation during apoptosis. *Nature Rev. Mol. Cell Biol.* **5**, 897–907 (2004).
- Wang, X. The expanding role of mitochondria in apoptosis. *Genes Dev.* **15**, 2922–2933 (2001).
- Li, P. *et al.* Cytochrome *c* and dATP-dependent formation of Apaf-1/Caspase-9 complex initiates an apoptotic protease cascade. *Cell* **91**, 479–489 (1997).
- Zou, H., Li, Y., Liu, X. & Wang, X. An APAF-1-cytochrome *c* multimeric complex is a functional apoptosome that activates procaspase-9. *J. Biol. Chem.* **274**, 11549–11556 (1999).
- Jiang, X. & Wang, X. Cytochrome *c* promotes caspase-9 activation by inducing nucleotide binding to Apaf-1. *J. Biol. Chem.* **275**, 31199–31203 (2000).
- Hu, Y., Benedict, M. A., Ding, L. & Nunez, G. Role of cytochrome *c* and dATP/ATP hydrolysis in Apaf-1-mediated caspase-9 activation and apoptosis. *EMBO J.* **18**, 3586–3595 (1999).
- Saleh, A., Srinivasula, S. M., Acharya, S., Fishel, R. & Alnemri, E. S. Cytochrome *c* and dATP-mediated oligomerization of Apaf-1 is a prerequisite for procaspase-9 activation. *J. Biol. Chem.* **274**, 17941–17945 (1999).
- Rodriguez, J. & Lazebnik, Y. Caspase-9 and Apaf-1 form an active holoenzyme. *Genes Dev.* **13**, 3179–3184 (1999).
- Inohara, N. & Nunez, G. The NOD: a signaling module that regulates apoptosis and host defense against pathogens. *Oncogene* **20**, 6473–6481 (2001).
- Hu, Y., Ding, L., Spencer, D. M. & Nunez, G. WD-40 repeat region regulates Apaf-1 self-association and procaspase-9 activation. *J. Biol. Chem.* **273**, 33489–33494 (1998).
- Srinivasula, S. M., Ahmad, M., Fernandes-Alnemri, T. & Alnemri, E. S. Autoactivation of procaspase-9 by Apaf-1-mediated oligomerization. *Mol. Cell* **1**, 949–957 (1998).
- Kaufmann, E. & Knöchel, W. Five years on the wings of fork head. *Mech. Dev.* **57**, 3–20 (1996).
- Holm, L. & Sander, C. Protein structure comparison by alignment of distance matrices. *J. Mol. Biol.* **233**, 123–138 (1993).
- Lenzen, C. U., Steinmann, D., Whiteheart, S. W. & Weis, W. I. Crystal structure of the hexamerization domain of N-ethylmaleimide-sensitive fusion protein. *Cell* **94**, 525–536 (1998).
- Zhang, X. *et al.* Structure of the AAA ATPase p97. *Mol. Cell* **6**, 1473–1484 (2000).
- Lupas, A. N. & Martin, J. AAA proteins. *Curr. Opin. Struct. Biol.* **12**, 746–753 (2002).
- Jaroszewski, L., Rychlewski, L., Reed, J. C. & Godzik, A. ATP-activated oligomerization as a mechanism for apoptosis regulation: fold and mechanism prediction for CED-4. *Proteins* **39**, 197–203 (2000).

18. Qin, H. *et al.* Structural basis of procaspase-9 recruitment by the apoptotic protease-activating factor 1. *Nature* **399**, 547–555 (1999).
19. Genini, D. *et al.* Nucleotide requirements for the *in vitro* activation of the apoptosis protein-activating factor-1-mediated caspase pathway. *J. Biol. Chem.* **275**, 29–34 (2000).
20. Leoni, L. M. *et al.* Induction of an apoptotic program in cell-free extracts by 2-chloro-2'-deoxyadenosine 5'-triphosphate and cytochrome c. *Proc. Natl Acad. Sci. USA* **95**, 9567–9571 (1998).
21. Cain, K., Brown, D. G., Langlais, C. & Cohen, G. M. Caspase activation involves the formation of the aposome, a large (~700 kDa) caspase-activating complex. *J. Biol. Chem.* **274**, 22686–22692 (1999).
22. Bochtler, M. *et al.* The structures of HsIU and the ATP-dependent protease HsIU-HsIV. *Nature* **403**, 800–805 (2000).
23. Gai, D., Zhao, R., Li, D., Finkielstein, C. V. & Chen, X. S. Mechanisms of conformational change for a replicative hexameric helicase of SV40 large tumor antigen. *Cell* **119**, 47–60 (2004).
24. Acehan, D. *et al.* Three-dimensional structure of the apoptosome: Implications for assembly, procaspase-9 binding and activation. *Mol. Cell* **9**, 423–432 (2002).
25. Otwinowski, Z. & Minor, W. Processing of X-ray diffraction data collected in oscillation mode. *Methods Enzymol.* **276**, 307–326 (1997).
26. Terwilliger, T. C. & Berendzen, J. Automated structure solution for MIR and MAD. *Acta Crystallogr. D* **55**, 849–861 (1999).
27. Jones, T. A., Zou, J.-Y., Cowan, S. W. & Kjeldgaard, M. Improved methods for building protein models in electron density maps and the location of errors in these models. *Acta Crystallogr. A* **47**, 110–119 (1991).
28. Collaborative Computational Project No 4. The CCP4 suite: programs for protein crystallography. *Acta Crystallogr. D* **50**, 760–763 (1994).
29. Kraulis, P. J. Molscript: a program to produce both detailed and schematic plots of protein structures. *J. Appl. Crystallogr.* **24**, 946–950 (1991).
30. Nicholls, A., Sharp, K. A. & Honig, B. Protein folding and association: insights from the interfacial and thermodynamic properties of hydrocarbons. *Proteins Struct. Funct. Genet.* **11**, 281–296 (1991).

Supplementary Information accompanies the paper on www.nature.com/nature.

Acknowledgements We thank M. Becker, A. Saxena, K. Dedrick and L. Walsh for assistance, A. Wist for help with the TLC, L. Gu for help with the omit map, F. Hughson for critically reading the manuscript, and members of the Shi laboratory for discussion. This work was supported by the NIH. S.J.R. is a Fellow of the Leukemia and Lymphoma Society.

Competing interests statement The authors declare that they have no competing financial interests.

Correspondence and requests for materials should be addressed to Y.S. (yshi@molbio.princeton.edu). The atomic coordinates have been deposited with the Protein Data Bank with the accession number 1Z6T.

Contacts

Publisher: Ben Crowe
Editor: Paul Smaglik
Marketing Manager: David Bowen

US Head Office, New York

345 Park Avenue South, 10th Floor,
New York, NY 10010-1707
Tel +1 800 989 7718
Fax +1 800 989 7103
e-mail: naturejobs@natureny.com

US Sales Manager/ Corporations:

Peter Bless

Classified Sales Representatives:

Phone: +1 800 989 7718

New York/ Pennsylvania/

Latin America:

Kelly Roman

Midwest USA/ Maryland/ NIH:

Wade Tucker

East USA/ Canada:

Janine Taormina

San Francisco Office

Classified Sales Representative:

Michaela Bjorkman

West USA/ West Corp. Canada

225 Bush Street, Suite 1453

San Francisco, CA 94104

Tel +1 415 781 3803

Fax +1 415 781 3805

e-mail: m.bjorkman@nature.sf.com

European Head Office, London

The Macmillan Building, 4 Crinan Street,
London N1 9XW, UK

Tel +44 (0) 20 7843 4961

Fax +44 (0) 20 7843 4996

e-mail: naturejobs@nature.com

Naturejobs Sales Director:

Nevin Bayoumi (4978)

European Sales Manager:

Andy Douglas (4975)

Advertising Production

Manager:

Billie Franklin

To send materials use London

address above.

Tel +44 (0) 20 7843 4814

Fax +44 (0) 20 7843 4996

e-mail: naturejobs@nature.com

Naturejobs web development:

Tom Hancock

Naturejobs online production:

Niamh Shields

European Satellite Office

Germany/ Austria/ Italy/

The Netherlands/ Belgium:

Patrick Phelan

e-mail: p.phelan@nature.com

Reya Silao

e-mail: r.silao@nature.com

Japan Head Office, Tokyo

MG Ichigaya Building (5F),

19-1 Haraikatamachi,

Shinjuku-ku,

Tokyo 162-0841

Tel +81 3 3267 8751

Fax +81 3 3267 8746

Asia-Pacific Sales Director:

Rinoko Asami

e-mail: rasami@nature.jp

naturejobs

Out in the cold

Maybe Larry Summers is secretly trying to do women scientists a big favour. Perhaps the embattled Harvard president's remarks about differences in "intrinsic aptitude" between men and women scientists (see *Nature* 433, 790; 2005) were actually a clever publicity ploy for Harvard, and a covert way to advance women scientists' issues — the latter of which can be a sisyphian struggle, at best.

Over the past few decades, with clock-like regularity, various scientific societies, non-governmental organizations and federal agencies have trotted out the same dismal statistics about women scientists. They are under-represented, paid less than men and tend to hit a glass ceiling. Various officials would express alarm for a week or two, promising progress, then the publicity would fade, and the stone would inch back down the slope.

But now, after Summers' remarks, every report about women in science will be subject to almost as much scrutiny as Summers himself. The most recent US report, by the American Institute of Physics (AIP), paints a mixed picture at best. It says that the percentage of women holding faculty positions in physics and astronomy is consistent with the percentage of women who earned degrees in the past — but the past hasn't been stellar. More women are entering science than before, but fewer of them are going into physics compared with some other fields. And the usual problems — including salary differences and the dearth of women from ethnic minorities in physics — persist.

But perhaps the spotlight Summers has turned on women and science will help to change things — even if a summary of the report by the AIP offers a subtle rebuke to Summers' unfortunate wording of the cause: "Women's participation in physics varies historically and across countries, pointing to the influence of cultural factors on women's rate of participation in science."

Paul Smaglik
Naturejobs editor



Contents

CAREERS AND RECRUITMENT

Alternative energy
hots up p936

CAREER VIEW

Recruiters & Industry
Microsoft's European
perspective
Graduate Journal
Home truths
Movers
Carlo Croce p938

WWW.NATUREJOBS.COM

Career centre
Information on the
scientific job market

FOCUS

SPOTLIGHT

RECRUITMENT

ANNOUNCEMENTS

EVENTS

High-energy career lines

As supplies of fossil fuels dwindle, the world is searching for alternative energy supplies. Materials scientists are in demand, says Virginia Gewin, but there are jobs in many areas.

Big oil companies are so last century. Shell is now a group of companies that includes Shell Renewables and its components Shell Solar and Shell WindEnergy. Lest any stone go unturned, the company is also on the trail of hydrogen, geothermal energy and biofuels. No one source of energy can meet the demands of the future.

The amount produced from renewable sources — particularly wind, sun and biofuels — is growing, but these are still costly. Fuel cells have promised more than they have yet delivered, and the infrastructure needs of hydrogen make that a long-term solution.

Although engineers and chemists of all kinds are needed, it is materials scientists who are in huge demand. Short of government funding, academia is relying on partnerships with industry — which, fortunately, has plenty of cash for training.

“The United States has shifted to long-range, high-risk research, whereas Europe is sticking with technologies in place today,” says Rick Sellers, head of the International Energy Agency’s Renewable Energy Unit.

The European Union (EU) aims to obtain 20% of its energy from renewable sources by 2020. Backed by development-friendly policies, this has created a robust wind and solar industry. The renewables community received €810 million (US\$1 billion) under the EU’s Sixth Framework Programme for funding research and



technology, which ends in 2006. They hope that the seventh, running till 2010, will double this amount.

The executive body, the European Commission (EC), tends to fund big industry-linked projects. “Research in industry is going to be more important in the coming years,” says Arthouros Zervos, president of the European Renewable Energy Council in Brussels. Industry provided 34% of funding at the Fraunhofer Institute for Solar Energy Systems in Freiburg, Germany, last year. The institute is a good stepping-stone to an industry career, says its spokeswoman, Karin Schneider.

REVIVAL OF ENERGY

The United Kingdom is reviving government spending on energy research with a number of new initiatives, notably a £28-million (US\$52-million) programme called Towards a Sustainable Energy Economy. A new body, the UK Energy Research Centre (UKERC), will coordinate efforts such as the Sustainable Power Generation and Supply project, or Supergen. Created by the Engineering and Physical Sciences Research Council (EPSRC), this will focus on four areas: marine, biomass and biofuels, hydrogen, and ‘future network technologies’ — taking account of possible trends such as widespread use of small-scale generators.

The UKERC will also support training programmes such as the Interdisciplinary Postgraduate Energy Research Studentships. Marie Curie fellowships, which provide European placements through the Framework programmes, are a staple source of postdoc funding. Other flexible training opportunities are also under way. In Britain, the EPSRC-funded Renewable Energy Flexible Training Programme provides paid full- and part-time studentships that mix web-based distance learning with internships. The new European master’s

Hot opportunities in Iceland

Iceland is already a leader in renewable energy usage, meeting 72% of its needs from its geothermal and hydropower resources. But transportation remains the obstacle to its goal to wean itself off fossil fuels. The transition has already begun with three fuel-cell buses



(right) on the streets — and the world’s first hydrogen filling station.

Future efforts are focused on meeting the need for an estimated 100,000 tonnes of hydrogen each year. Hydrogen storage remains the problem, says Ingolfur Thorbjornsson, a materials scientist at the Technological Institute of Iceland. He adds that Iceland has become the hot area for both Icelandic and foreign students to study renewables and the transition to hydrogen. **V.G.**



High winds: an engineer carries out essential maintenance work on turbines at a wind farm, one of the most established sources of renewable energy.

degree in renewable energy, sponsored by the EU Renewable Energy Centres agency, is aimed at engineers who want to specialize in any renewable technologies, and is available at seven universities.

Although training is flourishing in Europe, it remains flat in the United States. In 2005, the US Department of Energy will spend just over \$1 billion on renewables research. Most of that stays at the national labs, with some trickling down to support students at nearby universities. But funding at US universities is bleak, says Stephen Connors, research coordinator at the Massachusetts Institute of Technology's Laboratory for Energy and the Environment.

The bulk of US research is done at the National Renewable Energy Laboratory in Golden, Colorado. This hosts some 45 postdocs, plus 30–40 undergraduate and graduate students on the Research Participant Program, says Beverly Maestas, director of human resources.

The Petroleum Research Fund, established by seven oil companies in 1944 and now administered by the American Chemical Society, offers postdoctoral alternative-energy fellowships. But these may soon fall victim to funding shortfalls, says programme manager Ron Siatkowski.

Industry also created the \$225-million, ten-year Global Climate and Energy Project (GCEP) at Stanford University, arguably the largest university-based energy-research programme in the United States.

"The goal of the GCEP is to build a group of students that will be needed to transform the world's energy system," says director Lynn Orr. "If we guess now what the shape of energy systems will be in 2050, we are certain to get it wrong." Fundamental research at universities can investigate many ideas, he adds,

"Fuel-cell companies have had a tough time coming up with products to sell at a profit."

— Nick Lenssen

that may or may not be developed at a large scale.

"We'd love it if there were a dozen more programmes like ours to interact and work with," says Orr — one reason why the GCEP is making partnerships with institutions such as the Energy Research Centre of the Netherlands, Delft University of Technology and the Swiss Federal Institute of Technology in Zurich.

The Link Foundation, which fosters education and innovation, is the only US non-governmental body without industry ties to offer PhD fellowships. It provides two-year scholarships worth \$25,000 a year.

THE HARD CELL FOR HYDROGEN

One area in which industry has scaled back R&D is fuel cells. "Fuel-cell companies have had a tough time coming up with products to sell at a profit, causing the venture-capital community to view them as a high-risk investment," says Nick Lenssen, senior director of market-intelligence company Primem in Boulder, Colorado. Governments, though, are backing the technology. Iceland has the stated goal of becoming the world's first hydrogen economy (see 'Hot opportunities in Iceland', opposite).

The US Solid State Energy Conversion Alliance is halfway through a ten-year partnership of government, industry and academia to create a cost-effective mass-produced 3–10-kilowatt solid-oxide fuel cell.

"We've set up satellite centres at several universities," says Gary McVay, manager of the materials-science department at the Pacific Northwest National Laboratory in Richmond, Washington. This way, universities can build expertise and have students able to run the new energy-conversion devices. The big hurdle is materials research. They need people with expertise in materials science, metallurgy and ceramic engineering, says McVay.

The \$9.4-million Texas Center for Superconductivity and Advanced Materials at the University of Houston is another partner, doing cutting-edge research. And the National Science Foundation has entered energy funding with an Industry/University Cooperative Research Center for Fuel Cells at the University of South Carolina, supporting 35 graduate students.

In Europe, the EC's Hydrogen and Fuel Cell Technology Platform continues to fund basic research. Individual countries, such as Germany, have reduced previous levels of funding, so German research institutions rely more on the EC, says Robert Steinberger-Wilckens, a fuel-cell researcher at the Research Centre Jülich. But the Fuel Cell Education and Training Center in Ulm is a bright spot, currently setting up new training opportunities.

Others see great opportunities for fuel cells. New applications keep generating new companies, says Werner Tillmetz, a researcher at the Centre for Solar Energy and Hydrogen Research in Baden-Württemberg, who hopes these will energize the industry.

Europe has been slow to invest in new companies, especially those promoting novel technology, says Phil Doran of venture-capital firm Core Technology Ventures in Schmitten, Germany. But he is convinced that fuel-cell technology in the EU merits investment. Perhaps his firm will 'strike oil' in a time when no one knows what will supply the energy of the future. ■

Virginia Gewin is a freelance science writer based in Portland, Oregon.

Web links

UKERC studentships

► www.ukerc.ac.uk/2005/content/view/48/59

University of Newcastle upon Tyne's Reflex programme

► www.ncl.ac.uk/postgraduate/taught/subjects/martech/courses/418

European Master's in renewable energy

► www.eurec.be/REMaster_intro.htm

EU Hydrogen and Fuel Cell Technology Platform

► www.hfpeurope.org

Global Climate and Energy Program at Stanford University

► gcep.stanford.edu
Fuel Cell Education and Training Center, Ulm

► www.fuelcell-educationcenter.com

US DOE Research Participant Program

► www.nrel.gov/hr/employment/rpp

US DOE National Renewable Energy Lab

► www.nrel.gov

GRADUATE JOURNAL

Home truths

I am amazed at the way Europe is increasingly becoming one big 'village'. The countries seem to be getting closer to each other. Now that Poland is a member of the European Union, for example, we can travel without restrictions. Not so long ago, it was very hard to leave Poland for Western Europe or the United States — especially to work or study.

I remember the problems faced by my father when he wanted to work abroad. My generation has incredible freedom by comparison: nearly all of my colleagues from my days as an undergraduate are abroad doing PhDs, or have simply moved away with their spouses or partners.

On the one hand, this is good — I now have friends in many different countries and can visit them whenever I want. But I do wonder what the future holds for Polish science. I understand that there are better opportunities for young scientists outside Poland, but will these well-trained scientists ever return home to work?

I hope that the answer is yes, although I have my doubts. A few of my colleagues and I rejected PhD posts abroad and stayed here, partly because we wanted to prove that it is possible to become a good scientist in Poland. What effect this choice will have on our careers remains to be seen. But at least we did what our hearts dictated. ■

Karolina Tkaczuk is a graduate student at the Technical University of Lodz, Poland.

RECRUITERS & INDUSTRY

Microsoft's European perspective

As part of its European Science Initiative, software giant Microsoft has announced that it intends to open up to five 'centres of excellence' in Europe over the next five years. The company is also funding a career-development fellowship programme, which will see up to ten European postdocs in science and technology receive as much as €250,000 (US\$320,000) each. In addition, Microsoft will support up to 30 PhDs a year, fund scientific workshops throughout the continent and sponsor an award for European Scientist of the Year.

All of these efforts will help Microsoft to find ways to match its computing expertise with different scientific disciplines in Europe, says Stephen Emmott, director of the company's External

Research Office in Cambridge, UK. It also aims to match its own research expertise in computing with promising European scientists in emerging fields that embody "new kinds of science", he adds.

To that end, the first centre of excellence will be the Centre for Computational and Systems Biology at the University of Trento in Italy. The university and Italian local and national governments are also providing funding. The centre will employ ten full-time research directors, and with postdocs, technicians and visiting scientists, should play host to about 40 scientists at any one time.

Microsoft hope that the five centres of excellence will help to build bridges between computing and other sciences, as well as between industry and academia, but it also wants them to serve as a focal

point for career development, says Emmott. The company aims to match promising young researchers with more senior scientists throughout Europe, as well as computer scientists at the company, and provide them with strong funding and infrastructure. "A key part is supporting the careers of various promising scientists," says Emmott.

For both graduate students and postdocs, the company is seeking young scientists willing to work at interfaces: between the United States and Europe; the company and academia; and computing and more traditional disciplines.

The other centres of excellence have yet to be announced, but sites in France, Germany and Britain are currently under consideration. ■

Paul Smaglik is editor of Naturejobs.
♦ <http://research.microsoft.com/ero>

MOVERS Carlo Croce, director, cancer genetics programme, Ohio State University



Carlo Croce lives for the rush from a discovery, which might explain why he can't think of anything else he would rather do with his life. His obsession is a cure for cancer. His career, always focused on that singular goal, has left dozens of discoveries in its wake.

While at medical school in his native Italy, Croce turned his growing interest in oncogenic viruses into a fellowship to study in the United States. But a freak

event — his mentor Karl Habel's contraction of encephalitis from a monkey B virus infection — diverted his fellowship from the Scripps Clinic in San Diego, California, to the Wistar Institute in Philadelphia, Pennsylvania. There, he mapped the first viral integration site on a chromosome — and started down the uncharted road of cancer genetics.

Within two years of coming to the United States, Croce had built a lab at the Wistar and in the early 1980s, he delivered dozens of high-profile papers illuminating the role of genetics in cancer. "Nobody was thinking cancer was a genetic disease," he says.

Croce went on to detail the role played by genes in lymphomas and tumours, and he recently defined the part that microRNAs play in cancer development.

"Essentially, all of my work is to try to identify the earliest genetic changes that occur in many different human cancers,"

Croce says. His major challenge has been to exploit the results from fundamental science to develop novel therapies.

"Without fundamental discoveries, we'll never cure cancer," he says.

Building an environment to churn out such discoveries — which prompted his move to Ohio State University — is his latest endeavour. Besides his new role as the director of the university's human cancer genetics programme, he is also developing an Institute of Genetics. "I want the university to be a world leader of cancer research," he says. And he wants its discoveries translated into treatments.

By August 2006, Croce will have more than 30 faculty members in his new building. A risk-taker at heart, he anticipates that the centre will allow him to continue scoring monumental discoveries. "You get high only if you make a really exciting discovery, the usual kind of stuff doesn't work," he says. Spoken like a true addict. ■

CV **2005:** Director, human cancer genetics programme, Ohio State University.

1991–2004: Director, Kimmel Cancer Institute, Philadelphia, Pennsylvania.

1988–91: Director, Fels Institute for Cancer Research and Molecular Biology, Temple University, Philadelphia, Pennsylvania.

1980–88: Professor of human genetics, University of Pennsylvania, Philadelphia, Pennsylvania.

1980–88: Associate director, the Wistar Institute, Philadelphia, Pennsylvania.

I love liver: a romance

A design for life.

Larissa Lai

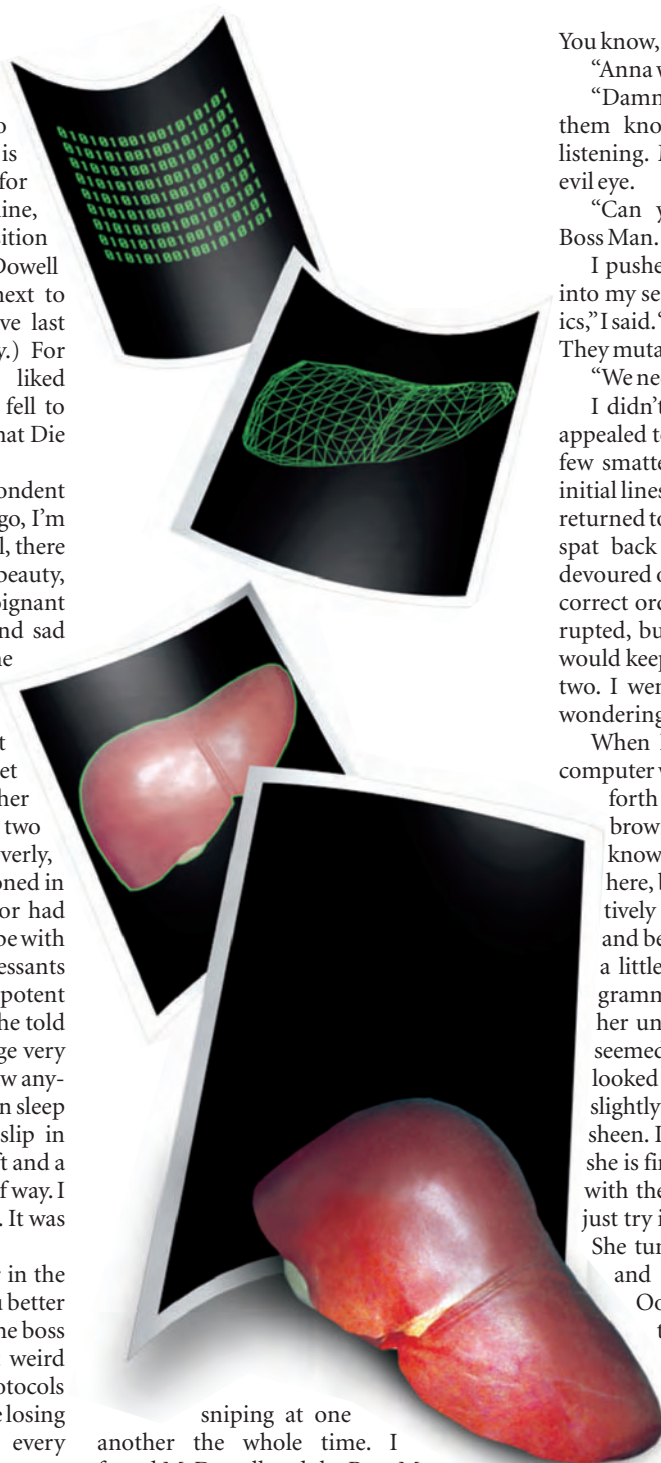
It has taken me almost four weeks of late nights and taxed my mochaccino machine to the limits. But Mira is ready. She began as a prototype for FreshCleanse's Liver Replacement line, but her capacity for toxin decomposition was weaker than that of the liver McDowell Hill came up with in the cubicle next to mine. (What's with people who have last names for first names? It's so tacky.) For whatever reason, the Boss Man liked Mackie's design better, and so Mira fell to the waste heap of Great Inventions That Die on the Drawing Board.

To be honest, I felt quite despondent about it. It wasn't just a blow to the ego, I'm used to those. It was more that ... well, there was something about Mira, a kind of beauty, extraordinary really. Something poignant about her lines, something tender and sad about her soft, brown-grey texture. The fact that she would never go into production threw me into a bit of a funk.

It took me a few days to realize that this wasn't something I would just get over, as I have with countless other designs. By the fourth day, even after two mochaccinos and a double dose of Beverly, my despondency seemed worse. I phoned in sick and went back to bed. My doctor had expressly told me how careful I had to be with Beverly. "This generation of antidepressants is more precise but also much more potent than what you're probably used to," she told me, "so you have to watch your dosage very carefully." Whatever. It was too late now anyway. I closed my eyes. Halfway between sleep and waking, I thought I saw Mira slip in beside me, larger than life, pillowy soft and a little slippery, in a smooth, sleek sort of way. I reached out to caress an elegant fluke. It was almost comforting.

A ringing phone woke me at four in the afternoon. It was McDowell Hill. "You better get down here right away," he said. "The boss doesn't care how sick you are. That weird liver you designed — it's jumped protocols and has infected the mainframe. We're losing thousands of hours of R&D with every minute that passes. You better get your pathetic, depressed butt down here ASAP." For a minute I thought: 'Who cares, you smarmy creep. I hope Mira burns the whole operation down.' But then I'd be out of a job. I got my pathetic, depressed butt down to the office.

The place was in chaos. The Tech Support boys were mousing as fast as their caffeine-pumped little hands could move, jibing and



sniping at one another the whole time. I found McDowell and the Boss Man at my cubicle, rifling furiously through my password-protected files with brazen impunity. Who needs this job? I thought. "I don't understand how a liver design can go viral like that," the Boss Man was saying.

"Over-rationalization," said Mackie. "The protocols are too close. And that was one weird little liver Anna designed. That's what you get for hiring these foreigners.

You know, I'm not sure she's entirely stable."

"Anna was born here," the Boss Man said.

"Damn right," I said, by way of letting them know I'd been there behind them listening. Mackie turned, and shot me the evil eye.

"Can you fix this, Anna?" asked the Boss Man.

I pushed Mackie out of the way and slid into my seat. "That's the thing about organics," I said. "They aren't static. They do things. They mutate."

"We need better firewalls," Mackie said.

I didn't fix anything. It was more like, I appealed to Mira. I coaxed her gently with a few smatterings of code. I showed her the initial lines of a heart I was working on. Mira returned to her original storage location. She spat back most of the information she'd devoured on her rampage. It wasn't all in the correct order, and some of it had been corrupted, but it was pretty much all there. It would keep Tech Support busy for a week or two. I went back home to my depression, wondering if Mira was depressed too.

When I got back to my apartment, my computer was on. Mira was floating back and forth across the screen like a pretty brown-grey fish in an aquarium. I don't know how she got from FreshCleanse to here, but I suppose such things are relatively easy these days. I opened her up and began the modifications. I made her a little larger. I cribbed some slug programming off a biologist's website to give her underside motility. To give her eyes seemed too strange somehow. Antennae looked better. I altered her coloration just slightly to give her an attractive iridescent sheen. It's taken me a few weeks, but now she is finally ready to print. What's wrong with the print function? Never mind, I'll just try it again. There we go. Hello, Mira! She tumbles gracefully from the printer and slithers across my office floor.

Oops. Must have pressed 'print' twice. Hello, Mira Two! ... Oh no, something is wrong. Another flap of liver emerges from the printer. She's cute. I can manage three. Here comes another. Am I in some kind of trouble?

I'll let it go to 12 before I call

Tech Support.

Larissa Lai is the author of two novels, Salt Fish Girl (Thomas Allen, 2002) and When Fox Is a Thousand (Press Gang, 1995; Arsenal Pulp Press, 2004). She has an MA in creative writing from the University of East Anglia and is currently completing a PhD in English at the University of Calgary.